

# Recent Advances in the Modelling and Analysis of Opinion Dynamics on Influence Networks

Brian D. O. Anderson<sup>1,2,3</sup>      Mengbin Ye<sup>1,4</sup>

<sup>1</sup>Research School of Engineering, Australian National University, Canberra 2601, Australia

<sup>2</sup>School of Automation, Hangzhou Dianzi University, Hangzhou 310000, China

<sup>3</sup>Data61-Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra 2601, Australia

<sup>4</sup>Faculty of Science and Engineering, Engineering and Technology Institute Groningen (ENTEG), University of Groningen, Groningen 9747 AG, The Netherlands

**Abstract:** A fundamental aspect of society is the exchange and discussion of opinions between individuals, occurring in situations as varied as company boardrooms, elementary school classrooms and online social media. After a very brief introduction to the established results of the most fundamental opinion dynamics models, which seek to mathematically capture observed social phenomena, a brief discussion follows on several recent themes pursued by the authors building on the fundamental ideas. In the first theme, we study the way an individual's self-confidence can develop through contributing to discussions on a sequence of topics, reaching a consensus in each case, where the consensus value to some degree reflects the contribution of that individual to the conclusion. During this process, the individuals in the network and the way they interact can change. The second theme introduces a novel discrete-time model of opinion dynamics to study how discrepancies between an individual's expressed and private opinions can arise due to stubbornness and a pressure to conform to a social norm. It is also shown that a few extremists can create “pluralistic ignorance”, where people believe there is majority support for a position but in fact the position is privately rejected by the majority. Last, we consider a group of individuals discussing a collection of logically related topics. In particular, we identify that for topics whose logical interdependencies take on a cascade structure, disagreement in opinions can occur if individuals have competing and/or heterogeneous views on how the topics are related, i.e., the logical interdependence structure varies between individuals.

**Keywords:** Opinion dynamics, social networks, influence networks, agent-based models, multi-agent systems, networked systems.

## 1 Introduction

In the broad area of social network analysis, the topic of “opinion dynamics” has received significant attention from the systems and control engineering community over the past decade. Opinion dynamics is the development and analysis of dynamical models that capture how individuals in a social network interact and exchange opinions; an individual's opinion may evolve over time as a result of learning the opinions of his or her neighbour. Many opinion dynamics models, including the most popular ones, are agent-based models where each individual is represented by an agent and the opinion of an individual on a topic is represented by a real value, evolving in time. The network of interactions between individuals is conveniently captured by a graph, where a node represents an individual whilst edges represent an interaction between two individuals.

In 1956, French Jr<sup>[1]</sup> introduced an agent-based model

of opinion dynamics to study how individuals exerted social power on each other during interactions in a network. The model has become known as the French-DeGroot (or simply DeGroot<sup>[2]</sup>) model and is the fundamental agent-based model of opinion dynamics which many subsequent works, including those discussed in this paper, build upon. The model assumes that each individuals' opinion (which is a real number) evolves over time as he or she integrates learned opinion values of his or her neighbours with the individual's own opinion using a weighted averaging process (modelled with a difference or differential equation) to capture the concept of social influence. (This has in fact led to the term “influence network” as a shorthand description of such a model.) Eventually, a consensus is reached on the opinion value, i.e., there is agreement across the opinions of all individuals, if the network satisfies some connectedness conditions. Experimental validations of the DeGroot model are reported in [3, 4].

A quite different approach to the DeGroot model, not analysed in this paper, but which has also provided great insight is the use of kinetic modelling for describing opinion dynamics<sup>[5–8]</sup>. In this approach, ideas of statistical mechanics and the kinetic theory of gases are used as the

Review  
Manuscript received October 5, 2018; accepted December 29, 2018;  
published online February 2, 2019  
Recommended by Associate Editor Min Wu  
© The Author(s) 2019

basis for formulating the models; as such the models are essentially restricted to large scale networks, rather than say a network comprised of a company's board of directors. Molecules and velocities, a common ingredient of the kinetic models, are replaced by individuals and opinions in the opinion dynamics models. The well-developed tools of statistical mechanics such as Boltzmann and Fokker-Planck equations have their analogues, and often analytic determination of limiting (time going to infinity) values can be determined. Evolution of opinions can be modelled through diffusion or modelling opinion exchange (and associated modification) between agents. (We note that diffusion models are examined in some detail in [9] also, distinguishing types such as information cascade, linear threshold and epidemic, and their use for influence maximization and information source detection is explored). It is important to note that through a series of developments, the models of [5–8] can be given greater sophistication through, e.g., inclusion of leadership attributes, the identification and use of highly connected agents, mechanisms for the creation and destruction of interactions, and the inclusion of stubborn agents. (Such refinements may well be mirrored also in certain refinements of the DeGroot model.)

Acknowledging that for larger and larger networks, agent-based models may be less and less appropriate, we nonetheless turn to examining some developments of the DeGroot model. Smaller networks are still of significant interest, as many small deliberative groups make important decisions, e.g., jury panels, government cabinets, and company board of directors. Beyond a model capturing simple consensus, variations of the DeGroot model have been proposed to investigate how different social phenomena may arise, often by adjusting the agent dynamics to capture some additional aspect involved in an individual's learning and assimilating of learned opinions; the aim is to better capture real world networks, where there is often a diverse range of opinions on a given topic. The Hegselmann-Krause model<sup>[10–15]</sup> captured homophily using bounded confidence, where an individual interacts only with those others who have similar opinions. Over time, individuals can become separated into clusters of disconnected subgraphs, where the final opinions are the same within each cluster, but, different between the clusters.

Polarisation, in which the network separates into two clusters of opposing opinions, has been heavily studied. The Altafini model used negative edge weights to introduce the idea of antagonistic interactions among individuals who may, for any number of reasons, dislike or mistrust each other<sup>[16–19]</sup>. If the network is “structurally balanced”<sup>[20]</sup> and satisfies appropriate connectivity conditions, the opinions can become polarised into two opposing clusters. Other models incorporating negative interactions include [21–23]. Polarisation has also been attributed to an individual's propensity for biased assimilation of information sources<sup>[24]</sup>.

The majority of the above mentioned models capture weak diversity<sup>[21, 23]</sup>, where there is no difference between opinions in the same cluster. At the same time, there has been a growing interest to study models which are able to capture strong diversity<sup>[21, 23]</sup>, which is frequently observed in the real world. In such scenarios, the opinions eventually converge to a configuration of persistent disagreement, with a diverse range of opinion values (and there may be clusters of opinions with similar, but not equal, values within a cluster). One is particularly interested in strong diversity in social networks that retain some form of connectivity over time; it is rare to see real-world networks with eventually completely disconnected subgroups, as arises in the Hegselmann-Krause models. One particular high-level question suggesting itself is: If social influence is acting to bring opinions closer together, then what other process that must be at work in connected networks to generate strong diversity?

Mäs et al.<sup>[21, 23]</sup> consider two features. The first is “social distancing”, in which individuals place a negative weight on opinion values which they consider are too different from their own; the key difference to the antagonistic weights in the Altafini model is that [21, 23] consider weight magnitudes which depend on differences in opinions, whereas the Altafini model assumes constant, or time-varying (but state-independent) negative weights. The second is to capture an individual's “desire to be unique”, where a state-dependent noise grows in magnitude as the individual's opinion grows closer to the average opinion of the network. Amelkin et al.<sup>[25]</sup> assumes an individual's susceptibility to interpersonal influence is dependent on the individual's current opinion; strong diversity can then arise, but only in a special case of the model. The Friedkin-Johnsen model shows that strong diversity may occur due to an individual's stubborn attachment (which can vary in level of intensity) to his or her initial opinion<sup>[26]</sup>. The Friedkin-Johnsen model is notable amongst existing opinion dynamics models in that it has been extensively verified via laboratory experiments for small networks<sup>[27–29]</sup> and in a quasi-field experiment for medium-sized networks<sup>[30]</sup>. An extension to capture the simultaneous discussion of multiple logically interdependent topics was introduced in [31], and used to analyse the US population's shifting opinions regarding the 2003 US-led invasion of Iraq<sup>[32]</sup>.

Rather than providing a broad survey of all existing opinion dynamics works (for which<sup>[33–35]</sup> are suitable), we present a narrower and more detailed focus, allowing more reflective discussions. In particular, this paper will introduce and summarise a set of very recent works which extend the DeGroot and Friedkin-Johnsen models in three different and significant directions. First, we review other existing work on the recently introduced DeGroot-Friedkin model<sup>[36]</sup> and summarise several advances on the analysis of the model made by [37–39]. The DeGroot-Friedkin model considers a social network that discusses a

sequence of topics, with each discussion occurring using the DeGroot model dynamics. An issue of significant interest is the evolution of individual social power, which is the amount of weight that an individual accords his or her own opinion during the discussion process. Evolution occurs when one discussion topic is finished and before another begins. According to the DeGroot-Friedkin model, an individual's social power changes at the end of a discussion of one topic depending on how much influence he or she has in determining the outcome of the discussion; as expected, an individual's social power increases or decreases as his or her influence on the discussion increases or decreases, respectively.

Second, we present a novel opinion dynamics model to examine how discrepancies in the expressed and private opinions of the same individual can arise. The fact that an individual can hold a private opinion different to the one he or she expresses within a social setting is well established in the social sciences<sup>[40–42]</sup>. It is perhaps a remarkable fact that up to now, almost all opinion dynamics models assume that each individual holds a single opinion per topic (some models assume an individual holds multiple opinions on multiple topics but each topic has only one opinion associated with it). The proposed model, termed the expressed and private opinion (EPO) model, assumes that each individual has a separate expressed and private opinion that evolve separately. The individual's private opinion evolves according to a modified Friedkin-Johnsen model, while his or her expressed opinion is distorted from his or her private opinion by a pressure to conform to the average expressed opinion (which represents a group standard or norm). We provide extensive literature support for the model, then review the convergence results and analysis of the limiting opinion distribution. In particular, we highlight several new and insightful conclusions and the associated interpretations in the sociological context. Furthermore, we use the model to revisit two classical works from social psychology: Asch's conformity experiments<sup>[40]</sup> and Prentice and Miller's field experimental data on pluralistic ignorance regarding the acceptance of alcohol drinking culture on the Princeton University campus<sup>[43]</sup>.

The third direction we study focuses on a network of individuals discussing multiple logically interdependent topics. As an illustration of logically interdependent topics, consider the following two statements: 1) gay marriage should be permitted, and 2) a person's sexual orientation is largely genetically inherited. It is clear that an individual is likely to see these two matters as logically related, so that the individual's opinion on one may not evolve independently of his or her opinion on the other because of an internal belief system. The term belief system is used to connote a set of topics and the logical connections an individual places between the topics<sup>[44]</sup>. When a group of individuals interact expressing opinions on logically interdependent topics, it may be that the input to the thinking process of one individual from the other

group members is consistent with that individual's internal belief system, or it may be inconsistent. Roughly speaking, consensus is more likely when there is consistency. In this part of the paper, we both present a recently developed model<sup>[31, 32]</sup> and also obtain conclusions on convergence of opinions to a steady state, including the question of whether convergence to a consensus actually occurs, if not for opinions on all topics, then at least for opinions on at least one.

The remainder of this paper is organised as follows. Section 2 provides an introduction to fundamental modelling of opinion dynamics and associated mathematical tools and results. Section 3 introduces the DeGroot-Friedkin model of social power evolution and presents a number of new results. Directions for future work are also commented upon. Section 4 introduces the novel EPO model and identifies a number of interesting phenomena that arises from study of the model, and records again directions for future work. The treatment of opinion dynamics given logically related topics is treated in Section 5. Lastly, conclusions are presented in Section 6.

## 2 Modelling of opinion dynamics

In this section, we provide the reader with a detailed introduction to two fundamental models of opinion dynamics. To begin, we establish the mathematical notation to be used in this paper. The  $n$ -column vector of all ones and zeros is given by  $\mathbf{1}_n$  and  $\mathbf{0}_n$ , respectively. The  $n \times n$  identity matrix is given by  $\mathbf{I}_n$ . The  $i$ -th canonical base unit vector of  $\mathbf{R}^n$  is denoted as  $\mathbf{e}_i$ , i.e.,  $\mathbf{e}_i \in \mathbf{R}^n$  has one in its  $i$ -th entry and zeros elsewhere.

We say that a matrix  $\mathbf{A}$  is nonnegative (respectively positive) if all of its entries  $a_{ij}$  are nonnegative (respectively positive). The matrix  $\mathbf{A}$  is denoted as being nonnegative and positive by  $\mathbf{A} \geq 0$  and  $\mathbf{A} > 0$ , respectively. A nonnegative matrix  $\mathbf{A} \in \mathbf{R}^{n \times m}$  is said to be row-substochastic (respectively row-stochastic) if, for all  $i = 1, \dots, n$ , there holds  $\sum_{j=1}^m a_{ij} \leq 1$  (respectively  $\sum_{j=1}^m a_{ij} = 1$ ). A matrix  $\mathbf{A}$  is said to be doubly stochastic if  $\sum_{j=1}^n a_{ij} = 1$  and  $\sum_{i=1}^n a_{ji} = 1$ . The spectral radius of a square matrix  $\mathbf{A} \in \mathbf{R}^{n \times n}$  is the largest modulus value of the eigenvalues of  $\mathbf{A}$ , and is denoted by  $\rho(\mathbf{A})$ . For a matrix  $\mathbf{A} \in \mathbf{R}^{n \times n}$ ,  $\lambda_i(\mathbf{A})$  denotes an eigenvalue of  $\mathbf{A}$ . A useful definition for a certain matrix property is now given.

**Definition 1. (Primitivity, [45])** A nonnegative square matrix  $\mathbf{A}$  is primitive if there exists  $k \in \mathbf{N}$  such that  $\mathbf{A}^k > 0$ .

### 2.1 Graph theory

In this subsection, we introduce graphs and graph theory. A graph is a powerful tool for modelling the network of interactions between a group of individuals, and at times this paper will use the term "network" and "graph" interchangeably.

For a given nonnegative matrix  $\mathbf{A} \in \mathbf{R}^{n \times n}$ , we asso-

ciate with it a graph  $\mathcal{G}[\mathbf{A}] = (\mathcal{V}, \mathcal{E}[\mathbf{A}], \mathbf{A})$ , where  $\mathcal{V} = \{v_1, \dots, v_n\}$  is the set of nodes of  $\mathcal{G}[\mathbf{A}]$  and in the context of this paper, each node represents an individual in a population of size  $n$ . An edge  $e_{ij} = (v_i, v_j)$  is in the set of ordered edges  $\mathcal{E}[\mathbf{A}] \subseteq \mathcal{V} \times \mathcal{V}$  if and only if  $a_{ji} > 0$ , where  $a_{ij}$  is the  $(i, j)$ -th entry of  $\mathbf{A}$ . A self-loop for node  $v_i$  exists if  $e_{ii} \in \mathcal{E}[\mathbf{A}]$ . The edge  $e_{ij}$  is said to be incoming with respect to  $v_j$  and outgoing with respect to  $v_i$ , and connotes that  $v_j$  learns of some information (typically an opinion value) from  $v_i$ . We do not assume that  $\mathbf{A} = \mathbf{A}^T$  in general, and thus  $\mathcal{G}[\mathbf{A}]$  is in general a directed graph. The neighbour set of  $v_i$  is defined as  $\mathcal{N}_i = \{v_j \in \mathcal{V} : (v_j, v_i) \in \mathcal{E}[\mathbf{A}]\}$ . A directed path is a sequence of edges of the form  $(v_{p_1}, v_{p_2}), (v_{p_2}, v_{p_3}), \dots$ , where  $v_{p_i} \in \mathcal{V}$  and  $e_{p_i p_{i+1}} \in \mathcal{E}$ . Node  $i$  is reachable from node  $j$  if there exists a directed path from  $v_j$  to  $v_i$ . Moreover, a graph  $\mathcal{G}[\mathbf{A}]$  is strongly connected if and only if there is a path from every node to every other node<sup>[46]</sup>. A graph  $\mathcal{G}[\mathbf{A}]$  is strongly connected if and only if  $\mathbf{A}$  is irreducible<sup>[46]</sup>, or equivalently, there does not exist a reordering of the nodes  $\mathcal{V}$  such that  $\mathbf{A}$  can be expressed as a block triangular matrix. A directed cycle is a directed path that starts and ends at the same vertex, and contains no repeated vertex except the initial (which is also the final) vertex. The length of a cycle is the number of edges in the cyclic path. A graph is aperiodic if the smallest integer  $k$  that divides the length of every cycle of the graph is  $k = 1$ <sup>[45]</sup>. Note that any graph with a self-loop is aperiodic. A result linking  $\mathcal{G}[\mathbf{A}]$  to the primitivity of  $\mathbf{A}$  is now given.

**Lemma 1.** ([45]) The graph  $\mathcal{G}[\mathbf{A}]$  is strongly connected and aperiodic if and only if  $\mathbf{A}$  is primitive.

From results on nonnegative matrices and, further, the Perron-Frobenius theorem<sup>[47]</sup>, we can establish the following result.

**Lemma 2. (Dominant eigenvectors)** For a strongly connected graph  $\mathcal{G}[\mathbf{A}]$  with row-stochastic  $\mathbf{A}$ , there are strictly positive left and right eigenvectors  $\mathbf{u}^T$  and  $\mathbf{1}_n$  of  $\mathbf{A}$  associated with the simple eigenvalue  $\lambda_1 = \rho(\mathbf{A}) = 1$ . With normalisation satisfying  $\mathbf{u}^T \mathbf{1}_n = 1$ , we call  $\mathbf{u}^T$  and  $\mathbf{1}_n$  the dominant left and right eigenvectors of  $\mathbf{A}$ , respectively.

## 2.2 DeGroot and Friedkin-Johnsen models

We are now in a position to introduce the DeGroot and Friedkin-Johnsen (which is a powerful generalisation of DeGroot) models. Consider a population of  $n$  individuals, whose interactions are modelled by a graph  $\mathcal{G}[\mathbf{A}]$ , discussing a single topic. Individual  $i$  has an opinion  $x_i(k) \in \mathbf{R}$ , at discrete time instants  $k = 0, 1, \dots$ , and according to the DeGroot model, evolves as

$$x_i(k+1) = \sum_{j=1}^n a_{ij} x_j(k) \quad (1)$$

where the influence weights  $a_{ij} \geq 0$  satisfy  $\sum_{j=1}^n a_{ij} = 1$ ; this immediately implies that  $\mathbf{A}$  is nonnegative and row-stochastic. In compact form, the opinions of all individuals, recorded as  $\mathbf{x} = [x_1, \dots, x_n]^T$  evolve as

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k). \quad (2)$$

Convergence of the model has been extensively studied, and of particular interest is convergence to a consensus of opinions, which occurs if, for all  $\mathbf{x}(0)$ , the system (2) converges to  $\mathbf{x}(\infty) = \beta \mathbf{1}_n$ ,  $\beta \in \mathbf{R}$ . The conditions for convergence of (2) are summarised succinctly in [33], with the following result detailing conditions for consensus on strongly connected  $\mathcal{G}[\mathbf{A}]$ .

**Lemma 3.** Suppose that  $\mathcal{G}[\mathbf{A}]$  is strongly connected. Then, (2) converges to a consensus,  $\lim_{k \rightarrow \infty} \mathbf{x}(k) = \beta \mathbf{1}_n$ ,  $\beta \in \mathbf{R}$ , if and only if  $\mathcal{G}[\mathbf{A}]$  is aperiodic. Moreover,  $\beta = \zeta^T \mathbf{x}(0)$ , where  $\zeta^T$  is the dominant left eigenvector of  $\mathbf{A}$ .

Before we move to consider the Friedkin-Johnsen model, we provide several comments on the DeGroot model. First, defining  $x_i$  as a real number (as opposed to requiring it to be, say  $+1$  or  $-1$ ) is useful in a broad range of applications scenarios, and one might define an interval  $[a, b]$ , with  $a, b \in \mathbf{R}$ , such that  $x_i = a$  and  $x_i = b$  represent the two extreme views of the opinion interval<sup>1</sup>, while values of  $x_i$  in between represent an individual  $i$  with views of varying conviction. For example, the social network may be discussing a topic which is subjective (for which no exact answer exists), e.g., “was the 2003 US-led invasion of Iraq justified?”<sup>[32]</sup>. Alternatively, one could consider an intellectual topic (provably true or false), e.g., “smoking tobacco damages your lungs”. Other social network models<sup>[9, 48–50]</sup>, such as the diffusion/threshold model, define  $x_i$  as a discrete variable, and may be more suitable for opinions that lead to actions, e.g., voting choices for a political election.

Second, we consider only strongly connected graphs in this paper, even though many results can be extended to weaker graph connectivity requirements. This is because the focus is to advance the models themselves to study new phenomena, and thus strong connectivity serves as a suitable and convenient assumption that can be relaxed for future work. Third, the constraint that  $\sum_{j=1}^n a_{ij} = 1$  and  $a_{ij} \geq 0$  implies that  $x_i(k+1)$  is a convex combination, or weighted average, of  $x_j(k)$ ,  $j = 1, \dots, n$ . It turns out that (1) has been extensively studied as an averaging algorithm with application to multi-agent consensus and coordination, including with time-varying  $a_{ij}(k)$ , see e.g., [51–54].

We conclude by introducing the Friedkin-Johnsen

<sup>1</sup>Well constructed opinion dynamics models (such as the DeGroot and Friedkin-Johnsen models) have the property that  $x_i(0) \in [a, b]$ ,  $\forall i \in \{1, \dots, n\} \Rightarrow x_i(k) \in [a, b]$ ,  $\forall i \in \{1, \dots, n\}$  and  $\forall k \geq 0$ . Two common intervals are  $[-1, 1]$  and  $[0, 1]$ .



model, which is discussed in further detail in [27, 31, 33, 55]. Again considering a population of  $n$  individuals interacting on  $\mathcal{G}[\mathbf{A}]$ , individual  $i$ 's opinion evolves as

$$x_i(k+1) = \lambda_i \sum_{j=1}^n a_{ij} x_j(k) + (1 - \lambda_i) x_i(0) \quad (3)$$

where the  $a_{ij}$  have the same constraints as detailed below (1), and  $\lambda_i \in [0, 1]$  is individual  $i$ 's susceptibility to influence ( $1 - \lambda_i$  is sometimes termed  $i$ 's stubbornness). Thus,  $1 - \lambda_i$  represents an individual's attachment to his or her initial opinion  $x_i(0)$ , and a measure of the unwillingness of individual  $i$  to accept new information. If  $\lambda_i = 1$ , then individual  $i$  is maximally susceptible to interpersonal influence, and we recover (1). On the other hand,  $\lambda_i = 0$  implies individual  $i$  is maximally closed to interpersonal influence (in the DeGroot model, this occurs in the special case where individual  $i$  has no neighbours). Accordingly, the compact opinion dynamical system is given by

$$\mathbf{x}(k+1) = \mathbf{\Lambda} \mathbf{A} \mathbf{x}(k) + (\mathbf{I}_n - \mathbf{\Lambda}) \mathbf{x}(0) \quad (4)$$

with  $\mathbf{\Lambda} = \text{diag}(\lambda_i)$  being the diagonal matrix of susceptibilities. Notice that if every individual is maximally susceptible, i.e.,  $\lambda_i = 1, \forall i \in \{1, \dots, n\}$ , then from (4) we recover (2); for the following result, we assume there is at least one individual  $i$  with  $\lambda_i < 1$ . For strongly connected networks, the following convergence result is available, summarised from [31].

**Lemma 4.** Suppose that  $\mathcal{G}[\mathbf{A}]$  is strongly connected, and that  $\exists i, j \in \{1, \dots, n\}$  such that  $\lambda_i, \lambda_j < 1$ . Then,  $\rho(\mathbf{I}_n - \mathbf{\Lambda} \mathbf{A}) < 1$  and (4) converges exponentially fast to

$$\lim_{k \rightarrow \infty} \mathbf{x}(k) \triangleq \mathbf{x}^* = \mathbf{V} \mathbf{x}(0). \quad (5)$$

The matrix  $\mathbf{V} \triangleq (\mathbf{I}_n - \mathbf{\Lambda} \mathbf{A})^{-1} (\mathbf{I}_n - \mathbf{\Lambda})$  is row-stochastic, and thus each entry of  $\mathbf{x}^*$  is a convex combination of  $\mathbf{x}(0)$ .

Further to this result, there is an interesting conclusion related to strong diversity, a concept which was discussed in the Introduction. The Friedkin-Johnsen model on strongly connected networks will in general yield strong diversity of the limiting opinions whenever  $\exists i, j \in \{1, \dots, n\}$  such that  $\lambda_i, \lambda_j < 1$  and  $x_i(0) \neq x_j(0)$ . If every individual has some stubbornness, i.e.,  $\lambda_i < 1, \forall i \in \{1, \dots, n\}$ , then for generic initial conditions  $\mathbf{x}(0)$ , there holds  $x_i^* \neq x_j^*$  for any  $i \neq j$ . In other words, for almost all initial conditions, the limiting opinions of the individuals in the social network display strong diversity.

**Remark 1.** The models discussed in this section, and those that will be introduced in latter parts of this paper, are all discrete-time models. Naturally, continuous-time counterparts to each model are either available or may be

proposed. In particular, the Abelson<sup>[56]</sup> and Taylor<sup>[57]</sup> models are the continuous-time counterparts to the DeGroot and Friedkin-Johnsen model. In many, but not all instances, the same phenomena that arise in the discrete-time model, mutatis mutandis, also arise in the continuous-time model. Thus, this paper will not consider continuous-time models, but the results covered in the subsequent parts of this paper certainly can be studied in continuous-time as future work. We expect that many of the analysis techniques found in the extensive multi-agent systems literature will be applicable for analysis of more complicated continuous-time opinion dynamics models, given their similarity to continuous-time multi-agent consensus algorithms, see e.g., [58, 59].

### 3 Evolution of social power

Suppose an individual is participating in discussion in a strongly connected network which covers a number of different issues (topics) sequentially, with the issues indexed by the issue sequence  $\mathcal{S} = \{0, 1, 2, \dots\}$ . Under the DeGroot model, each issue  $s \in \mathcal{S}$  is discussed through to consensus (because of the strongly connected network, see Lemma 3), then the next issue is discussed, and so on. Suppose that the individual perceives during this process that they have less and less impact on the outcome of each discussion. Consequently, and intuitively, they become less and less confident of their own opinion. (The converse situation of having more and more impact and rising confidence can also occur of course). This self-confidence has been termed social power<sup>[36]</sup>, with the reasons becoming apparent in the sequel, following formal introduction of the model.

We are thus interested in modelling how a person evaluates their influence on a discussion, and how the updating of this person's self-confidence affects discussion on the next topic. We treat these matters in sequence, and first introduce the DeGroot-Friedkin model<sup>[36]</sup>, before covering several new and major advances. The discussion of any one issue proceeds according to the DeGroot model (2). We first define  $a_{ii}$ , the  $i$ -th diagonal entry of  $\mathbf{A}$  as individual  $i$ 's self-confidence. We allow  $a_{ii}(s)$  to change in some way to be specified below, and for issue  $s \in \mathcal{S}$ , individual  $i$ 's opinion  $x_i(k, s)$  evolves for  $k = 0, 1, \dots$  as

$$x_i(k+1, s) = a_{ii}(s) x_i(k, s) + (1 - a_{ii}(s)) \sum_{j \neq i}^n c_{ij} x_j(k, s) \quad (6)$$

with  $a_{ij}(s) \triangleq (1 - a_{ii}(s)) c_{ij}$ . Here,  $c_{ij} \geq 0$  is independent of  $s$  and represents the relative trust individual  $i$  accords to individual  $j \neq i$  (we explain shortly why we refer to this as relative trust). With  $c_{ii} = 0$ , we further impose that  $\sum_{j=1}^n c_{ij} = 1$  for all  $i \in \{1, \dots, n\}$ . Thus, it is clear that as  $a_{ii}(s)$  evolves along the issue sequence  $s = 0, 1, \dots$  (in a manner we will describe in the sequel), there

continues to hold  $\sum_{j=1}^n a_{ij}(s) = 1$  for all  $i \in \{1, \dots, n\}$  and for all  $s \in \mathcal{S}$ . In other words, the opinion discussion for each topic  $s \in \mathcal{S}$  is

$$\mathbf{x}(k+1, s) = \mathbf{A}(s)\mathbf{x}(k, s) \quad (7)$$

with

$$\mathbf{A}(s) = \text{diag}(a_{ii}(s)) + (\mathbf{I}_n - \text{diag}(a_{ii}(s)))\mathbf{C} \quad (8)$$

row-stochastic, with the matrix  $\mathbf{C}$  formed from the relative trust entries  $c_{ij}$ . Thus, the opinion discussion for any issue  $s$  is modelled by the DeGroot process. The focus of the DeGroot-Friedkin model is to propose a systematic mechanism for updating  $a_{ii}(s)$ , and we address this in the next subsection.

### 3.1 Evolution by reflected self-appraisal

Consider an  $s \in \mathcal{S}$  and  $\mathcal{G}[\mathbf{C}]$  that is strongly connected. If  $a_{ii}(s) < 1$  for all  $i$  and  $\exists j : a_{jj}(s) > 0$ , then  $\mathcal{G}[\mathbf{A}(s)]$  with  $\mathbf{A}(s)$  defined in (8) is strongly connected and aperiodic<sup>[36, 38]</sup>. Thus,  $\mathbf{x}(s, k)$  reaches a consensus as  $k \rightarrow \infty$  as per Lemma 3. If instead  $\exists j : a_{jj}(s) = 1$  and  $a_{ii}(s) < 1$  for all  $i \neq j$ , then  $\mathcal{G}[\mathbf{A}(s)]$  is such that there is a path from node  $v_j$  to every other node  $v_i$ ,  $i \neq j$  and  $v_j$  has no incoming edges<sup>[36, 38]</sup> (in this case,  $\mathcal{G}[\mathbf{A}(s)]$  is not strongly connected), and standard consensus results establish that  $\lim_{k \rightarrow \infty} \mathbf{x}(s, k) = x_j(0)\mathbf{1}_n$ . In both cases (we shall establish in the sequel that for a large and reasonable set of initial  $a_{ii}(0)$ , these are the only two cases possible), we can write that

$$\lim_{k \rightarrow \infty} \mathbf{x}(k, s) = \boldsymbol{\zeta}^T(s)\mathbf{x}(0, s)\mathbf{1}_n = \sum_{i=1}^n \zeta_i(s)x_i(0, s)\mathbf{1}_n \quad (9)$$

where  $\boldsymbol{\zeta}^T(s)$  is the dominant left eigenvector of  $\mathbf{A}(s)$  if  $\mathcal{G}[\mathbf{A}(s)]$  is strongly connected and aperiodic and  $\boldsymbol{\zeta}^T(s) = \mathbf{e}_j$  in the latter case. From the fact that  $\sum_{i=1}^n \zeta_i(s) = 1$ , one can see that  $\zeta_i(s)$  captures the relative contribution, termed social power, of individual  $i$  to the discussion of topic  $s$ . The DeGroot-Friedkin model proposes that each individual updates his or her self-confidence  $a_{ii}$  using reflected self-appraisal at the end of each topic discussion. Formally, the update is

$$a_{ii}(s+1) = \zeta_i(s) \quad (10)$$

and then for the next topic  $s+1$ , the influence matrix  $\mathbf{A}(s+1)$  is determined by (8) but with  $s+1$  replacing  $s$ . This replacement indicates that for issue  $s+1$ , individual  $i$  weights his or her own opinion relative to the opinions of others by the same weight as his or her contribution to the consensus value in issue  $s$ . It also indicates that the nature of the interactions between individuals, discounting any self-weighting, is constant with  $s$ . If

individual 1 finds individual 2 twice as reliable as individual 3 for topic 0, that proportionality relationship will hold for all issues if  $\mathbf{C}(s)$  is independent of  $s$ . However, what does change is the overall weight individual 1 gives to all opinions other than his or her own, since in adjusting the self-weighting to be  $a_{ii}(s+1) = \zeta_i(s)$ , a compensating adjustment for weighting placed on others' opinions,  $a_{ij}(s+1) = (1 - \zeta_i(s))c_{ij}(s)$  is necessary to ensure that  $\sum_{j=1}^n a_{ij}(s+1)$  remains equal to 1.

One key task is to establish the properties of  $\zeta_i(s)$  along the sequence of topics  $s = 0, 1, 2, \dots$ . To do this, Jia et al.<sup>[36, 38]</sup> showed that if  $\mathcal{G}[\mathbf{C}]$  is strongly connected, and the initial conditions<sup>2</sup> satisfied a)  $\exists j : a_{jj}(0) > 0$  and  $a_{ii}(0) < 1, \forall i$  or b)  $\exists j : a_{jj}(0) = 1$  and  $a_{ii}(0) < 1, \forall i \neq j$ , then

$$\boldsymbol{\zeta}(s+1) = \mathbf{F}(\boldsymbol{\zeta}(s)) \quad (11)$$

where

$$\mathbf{F}(\boldsymbol{\zeta}) = \begin{cases} \mathbf{e}_i, & \text{if } \zeta_i = 1 \text{ for any } i \\ \alpha(\boldsymbol{\zeta}) \begin{bmatrix} \frac{\gamma_1}{1 - \zeta_1} \\ \vdots \\ \frac{\gamma_N}{1 - \zeta_N} \end{bmatrix}, & \text{otherwise} \end{cases} \quad (12)$$

with  $\alpha(\boldsymbol{\zeta}) = \frac{1}{\sum_{i=1}^n \frac{\gamma_i}{1 - \zeta_i}}$ . Here,  $\gamma_i$  is the  $i$ -th entry of the

dominant left eigenvector  $\boldsymbol{\gamma}^T$  of  $\mathbf{C}$ . It was also shown that the map  $\mathbf{F} : \Delta_n \mapsto \Delta_n$  is continuous<sup>[36]</sup> and smooth<sup>[38]</sup> on  $\Delta_n$ , and that  $\boldsymbol{\zeta}(s) \in \Delta_n$  for all  $s > 0$ .

### 3.2 Recent advances in analysis of the DeGroot-Friedkin model

Jia et al.<sup>[36, 38]</sup> established a number of results on the evolution of the social power vector  $\boldsymbol{\zeta}(s)$ ,  $s \geq 0$ . We summarise the key convergence results in Theorem 1, and then provide detailed discussions of various additional conclusions of interest. Before doing so, we define a topology class that has special convergence properties in the DeGroot-Friedkin model.

**Definition 2. (Star graph)** A strongly connected graph  $\mathcal{G}[\mathbf{C}]$  is called a star graph if and only if there exists a unique node  $v_i$ , called the centre node, from which every edge in  $\mathcal{E}[\mathbf{C}]$  is either incoming or outgoing with respect to  $v_i$ .

**Theorem 1.** Consider the system in (11), with  $n \geq 3$  individuals' relative interactions captured by the strongly connected  $\mathcal{G}[\mathbf{C}]$ . Suppose that the initial conditions satisfy

<sup>2</sup>The authors of [36] first established  $\mathbf{F}$  for initial conditions satisfying  $\sum_{i=1}^n a_{ii}(0) = 1$ . Our paper<sup>[38]</sup> showed  $\mathbf{F}$  also holds for the more general case stated in this paper.

fy<sup>3</sup>  $\exists j : a_{jj}(0) > 0$  and  $a_{ii}(0) < 1, \forall i$ . Then,

1) If  $\mathcal{G}[\mathbf{C}]$  is a star graph, whose centre node is assumed to be  $v_1$  without loss of generality, then  $\lim_{s \rightarrow \infty} \zeta(s) = e_1$ . Convergence while asymptotic is not exponentially fast. All other fixed points  $e_j, j \neq 1$  of  $\mathbf{F}$  are unstable.

2) If  $\mathcal{G}[\mathbf{C}]$  is not a star graph, then  $\lim_{s \rightarrow \infty} \zeta(s) = \zeta^*$  exponentially fast. In particular,  $\zeta^* \in \text{int}(\Delta_n)$  is the unique fixed point of the map  $\mathbf{F}$  in the set  $\tilde{\Delta}_n$ . All other fixed points  $e_j, j = 1, \dots, n$  of  $\mathbf{F}$  are unstable.

The original proof of convergence in [36] used LaSalle's invariance principle to establish asymptotic convergence for both star and non-star  $\mathcal{G}[\mathbf{C}]$ . Exponential convergence for non-star  $\mathcal{G}[\mathbf{C}]$  was first established in [38] using nonlinear contraction analysis and a set of specialised calculations tailored specifically to the functional form of  $\mathbf{F}$  in (12). In the same paper, exponential convergence for star graphs was ruled out. An alternative proof of exponential convergence using a generalised Lefschetz-Hopf result from differential topology, which simultaneously established the uniqueness of  $\zeta^*$ , was provided in [39] and had the virtue of not appealing to the specific functional form of (11), but requiring certain topological properties of some general update map  $\zeta(s+1) = \mathbf{F}(\zeta(s))$ .

### 3.2.1 Analysis of final social power

For non-star  $\mathcal{G}[\mathbf{C}]$ , a number of further conclusions can be drawn regarding the fixed point  $\zeta^*$ . First, for any  $i, j \in \{1, \dots, n\}$ , there holds  $\zeta_j^* > \zeta_i^*$  if and only if  $\gamma_j > \gamma_i$ , and  $\zeta_j^* = \zeta_i^*$  if and only if  $\gamma_j = \gamma_i$ <sup>[36]</sup>. Here,  $\gamma_i$  is the  $i$ -th entry of  $\gamma^T$ , the left dominant eigenvector of  $\mathcal{G}[\mathbf{C}]$ . Clearly, the ranking of the  $\gamma_i$  (termed eigenvector centrality in some disciplines), also determines the ranking of the individuals' final social powers.

Actually, more can be established from  $\gamma_i$ . In fact, when  $\mathcal{G}[\mathbf{C}]$  is not a star graph, we<sup>[38]</sup> can show that there holds for any  $i \in \{1, \dots, n\}$ ,

$$\zeta_i^* \leq \frac{\gamma_i}{1 - \gamma_i}. \quad (13)$$

Thus, we are able to upper bound the final social power of any individual in the network. For networks  $\mathcal{G}[\mathbf{C}]$  with  $\gamma_i \leq \frac{1}{3}$  for all  $i$ , it is also possible to compute a bound on the convergence rate, see [38].

An interesting result on the accumulation of social power in non-star graphs was presented in [36]. Define  $\bar{\zeta} = 1 - \alpha(\zeta^*) \in (0, 1)$ , where  $\alpha$  was given below (12). Then, if  $\bar{\zeta} < 0.5$ , there holds for any individual  $i$ , 1)  $\zeta_i^* > \gamma_i$  if  $\gamma_i > \bar{\zeta}$ , 2)  $\zeta_i^* < \gamma_i$  if  $\gamma_i < \bar{\zeta}$  and  $\zeta_i^* = \bar{\zeta}$  if  $\gamma_i = \bar{\zeta}$ . If  $\bar{\zeta} \geq 0.5$ , then there is a unique individual  $i$  with  $\zeta_i^* > \gamma_i$ , while all other individuals  $j \neq i$  have  $\zeta_j^* < \gamma_j$ . By viewing  $\bar{\zeta}$  as a threshold value, this result identifies those

individuals who accumulate more social power  $\zeta_i^*$  than their share of the centrality measure  $\gamma_i$ , and vice versa.

### 3.2.2 Dynamic relative interaction topology

The possibility of allowing more time-variation than that captured by the  $a_{ii}(s)$  is a natural consideration. Suppose that the group in question is a cabinet of ministers. Each week they might meet and regularly discuss diverse topics, e.g., relating to defence, social security and the economy. Because of the different expertise of the different ministers, it would be logical for Minister 1 to vary the relative weight  $c_{1j}, j = 2, 3$  he or she puts on the opinions of Ministers 2 and 3 in discussing topics of a different character. Further, the composition of such a group can change over time; friendships may be formed or broken. This leads to the consideration of  $\mathcal{G}[\mathbf{C}(s)] = \mathcal{G}[\mathbf{C}_{\sigma(s)}]$ , where  $\sigma(s)$  is a switching signal that captures the topic-varying nature of the relative interaction matrix  $\mathbf{C}(s)$ , and  $\sigma(s)$  is independent of  $\zeta(s), s \geq 0$ . Accommodating such time-variation, including the important specialisation of periodic time-variation, is difficult, but not impossible. This may appear surprising, given that the system is now

$$\zeta(s+1) = \mathbf{F}_{\sigma(s)}(\zeta(s)) \quad (14)$$

with  $\mathbf{F}_{\sigma(s)}$  defined similarly to that in (12) but with  $\gamma_i(s) = \gamma_{i, \sigma(s)}$  replacing  $\gamma_i$ , for all  $i \in \{1, \dots, n\}$ . Although (14) is a nonlinear switching discrete-time system, the nonlinear contraction analysis advanced in [38] for the original dynamics (11) proves apt at handling this. We establish that, for the system (14) with  $\mathcal{G}[\mathbf{C}(s)]$  being a strongly connected non-star graph for all  $s$  and initial conditions  $a_{ii}(0)$  as detailed in Theorem 1, there holds

$$\lim_{s \rightarrow \infty} \zeta(s) = \zeta^*(s) \quad (15)$$

where  $\zeta^*(s), s \geq 0$  is termed the "unique limiting trajectory" of (14) that is determined solely by the sequence of switching  $\mathbf{C}_{\sigma(s)}$ . A special case is when  $\mathbf{C}_{\sigma(s)}$  changes periodically, e.g., when a cabinet revisits the same set of issues every week; in such instances,  $\zeta^*(s)$  is a periodic trajectory. The ranking result and threshold result with  $\bar{\zeta}$  detailed in Section 3.2.1 have not been established for dynamic topology systems. However, and perhaps surprisingly, the upper bound result (13) and convergence rate result detailed below (13), can be established for networks with dynamic topology though with obvious adjustments to account for the fact that convergence occurs to the unique limiting trajectory  $\zeta^*(s)$  rather than a fixed point  $\zeta^*$ .

The key conclusion from study of dynamically changing relative interaction topology is that sequential opinion discussion removes initial social power/self-confidence exponentially fast. True social power/self-confidence evolving via reflected self-appraisal, obtained in the limit of the sequence of topic discussions, is dependent only on the sequence of topology structures, i.e., the distinct

<sup>3</sup>The case where  $\exists j : a_{jj}(0) = 1$  and  $a_{ii}(0) < 1, \forall i \neq j$  leads to trivial dynamics where  $\zeta(s) = e_j, \forall s > 0$  and we thus ignore this.

agent-to-agent interactions. This is clearly illustrated in the simulation result, displayed in Fig. 1. Fig. 1 shows a network of 6 individuals discussing topics on a periodically changing network. For the same network of individuals, we initialise the system with two different sets of initial conditions,  $\hat{a}_{ii}(0) \neq \tilde{a}_{ii}(0)$  for every  $i \in \{1, \dots, n\}$  (both sets of initial conditions satisfy the hypothesis in Theorem 1). For any individual  $i$ , his or her social power trajectory  $\zeta_i(s)$  converges to the unique, periodic trajectory  $\zeta_i^*(s)$ ,  $s \geq 0$  regardless of the initial self-confidence (dotted line for  $\hat{a}_{ii}(0)$  and solid line for  $\tilde{a}_{ii}(0)$ ). Moreover, it is clear that the convergence is exponential; by the 8th topic, both the dotted and solid trajectories have converged.

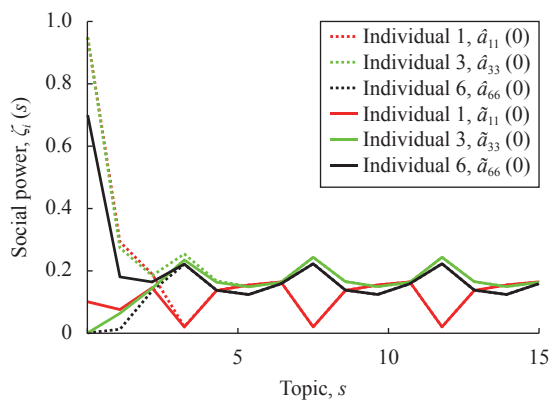


Fig. 1 Evolution of selected individuals' social powers over a sequence of topic discussions for a network of 6 individuals, with a periodically varying network structure. For each  $i = 1, 3, 6$ , the initial self-confidence  $\hat{a}_{ii}(0) \neq \tilde{a}_{ii}(0)$ . It can be seen that for any individual  $i$ , his or her social power trajectories from different initial conditions (dotted line for  $\hat{a}_{ii}(0)$  and solid line for  $\tilde{a}_{ii}(0)$ ) converge to the same unique trajectory  $\zeta_i^*(s)$ ,  $s \geq 0$ . (Color versions of the figures in this paper are available online.)

### 3.2.3 Similar time-scales, memory and noise

Now, we briefly touch upon other works which have advanced the original DeGroot-Friedkin model. First, the reader may have noticed that by assuming the self-appraisal dynamics follow (10), two restrictions are imposed on the social network. The first is that the time-scale for opinion evolution  $y_i(k, s)$  and social power evolution  $\zeta_i(s)$  are assumed to be separate. In particular, the opinion discussion occurs much faster than the self-appraisal, and thus a consensus is always reached before self-confidence  $a_{ii}(s)$  is updated. Second, (10) implies that the updating is centralised; any individual  $i$  knows precisely his or her relative contribution  $\zeta_i(s)$ . This is not a problem for small to medium sized networks, but may be less realistic for larger sized networks. A model which relaxes these two restrictions, first alluded to in [36], has been partially studied in [60, 61] as the "modified DeGroot-Friedkin model". A continuous-time counterpart has also been partially studied in [62]. A full analysis remains missing, though private communications with Professor

Bullo have indicated a more comprehensive result may be soon forthcoming.

Chen et al.<sup>[63]</sup> has provided a number of different advances to the original DeGroot-Friedkin model; we briefly summarise several contributions of note here. Switching topology with general  $\mathcal{G}[\mathbf{C}(s)]$  as in Section 3.2.2 is also considered, but with a major restriction that every  $\mathbf{C}(s)$  is a small perturbation from some fixed  $\mathbf{C}^*$ . Perhaps unsurprisingly given the conclusions in Section 3.2.2, it is established that  $\zeta(s)$  approaches a ball around the fixed point of  $\mathbf{F}$  with  $\mathbf{C}^*$ . Convergence for the model, modified to incorporate a sequence of stochastic  $\mathbf{C}(s)$ , environment noise, and in which there is "memory" in the entries of  $\mathbf{C}(s)$ , is also established. However, the conditions for convergence are complex and in general extremely difficult to verify.

## 3.3 Future research directions

We now describe some extensions of the above modelling currently under development. Interested readers may find a number of interesting problems for investigation.

### 3.3.1 Behaviour in self-appraisal dynamics

The first extension flows from the observation that during discussions, some individuals may be overconfident, other individuals may be underconfident, and other individuals again may over-react, either in a positive or negative direction to certain outcomes of the discussion.

As above,  $\zeta_i(s)$  is the social power or measure of the contribution individual  $i$  makes to the consensus value achieved for topic  $s$ . For topic  $s + 1$ , the individual's self-confidence  $a_{ii}(s + 1)$  updates via self-appraisal according to (10). However, in the extension we propose that during this self-appraisal process, an individual  $i$ 's personal behavioural characteristics may alter individual  $i$ 's perception of his or her social power  $\zeta_i(s)$ . This is captured using a function  $\phi_i(\zeta)$  that is assumed to have the following properties.

**Assumption 1.** For every  $i \in \{1, \dots, n\}$ ,  $\phi_i(x) : [0, 1] \rightarrow [0, 1]$  is a smooth monotonically increasing function satisfying  $\phi_i = 0 \Leftrightarrow x = 0$  and  $\phi_i = 1 \Leftrightarrow x = 1$ .

Then, we propose that the self-appraisal dynamics, for every individual  $i \in \{1, \dots, n\}$ , becomes

$$a_{ii}(s + 1) = \phi_i(\zeta_i). \quad (16)$$

An individual  $i$  whose was underconfident, or perhaps humble, might well have  $\phi_i(\zeta_i) < \zeta_i$  on  $(0, 1)$ , an overconfident or arrogant individual might have  $\phi_i(\zeta_i) > \zeta_i$  on  $(0, 1)$ , and an overly-reactive or emotional individual might have  $\phi_i(\zeta_i) < \zeta_i$  on  $(0, a)$  for some  $a \in (0, 1)$ ,  $\phi_i(a) = a$  and  $\phi_i(\zeta_i) > \zeta_i$  on  $(a, 1)$ . Finally, a balanced individual might have  $\phi_i(\zeta_i) = \zeta_i$  (i.e., the original model (10)). An example of an "emotional" individual is given in Fig. 2. This change means that the mapping  $\mathbf{F}$  from  $\zeta(s)$  to  $\zeta(s + 1)$  is modified to become



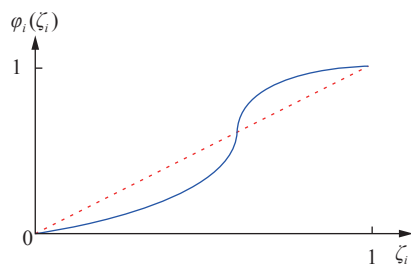


Fig. 2 Example of an emotional individual's  $\phi_i(\zeta_i)$ . The function  $\phi_i(\zeta_i)$  is plotted as the blue line, while the dotted red line corresponds to the original model,  $\phi_i(\zeta_i)$  being the identity mapping.

$$\bar{F} : \zeta(s) \rightarrow \zeta(s+1) = \frac{1}{\sum_{j=1}^n \frac{\gamma_j}{1 - \phi_j(\zeta_j(s))}} \begin{bmatrix} \frac{\gamma_1}{1 - \phi_1(\zeta_1(s))} \\ \vdots \\ \frac{\gamma_n}{1 - \phi_n(\zeta_n(s))} \end{bmatrix}. \quad (17)$$

Now it is the stability properties of this equation that need to be analysed. If all individuals are humble or balanced, there is as previously a single unique equilibrium in  $\text{int}(\Delta_n)$  which is approached exponentially fast under the same initial conditions as in Theorem 1. However, if one or more individuals are emotional, multiple attractive equilibria can exist in contrast to the original DeGroot-Friedkin model where for non-star graphs, there was a unique attractive equilibrium with all other equilibria being unstable. At this stage, only preliminary results<sup>[64]</sup> establishing the dynamical equations in (17) and convergence for some cases have been obtained.

### 3.3.2 Self-appraisal with stubborn individuals

The reader will likely have noticed that the DeGroot-Friedkin model assumes that during a discussion of any one topic, each individual's opinion evolves according to the DeGroot model as described in (1). Thus, under the suitable connectivity assumption of a strongly connected  $\mathcal{G}[\mathcal{C}]$ , a consensus is always achieved for each topic. However, we also introduced the Friedkin-Johnsen model in (3) and noted that it has been extensively validated in experiments. Thus, it is natural to consider the evolution of self-confidence with this model, i.e., the dynamics (10) but with (6) replaced by

$$x_i(k+1, s) = \lambda_i a_{ii}(s) x_i(k, s) + \lambda_i (1 - a_{ii}(s)) \sum_{j \neq i} c_{ij} x_j(k, s) + (1 - \lambda_i) x_i(0, s) \quad (18)$$

for every  $i \in \{1, \dots, n\}$  and  $s = 0, 1, \dots$ , and where  $\lambda_i \in [0, 1]$  is individual  $i$ 's susceptibility to influence as detailed below (3). In fact, the first reflected self-appraisal model proposed by Noah Friedkin in [65] considered (18) rather than (6). However, [65] primarily focused on introducing the idea of self-appraisal in

opinion dynamics along a sequence of topic discussions, with laboratory experiments and simulations. In fact, the DeGroot-Friedkin model with stubborn individuals has been extensively studied with simulations and validated empirically in e.g., [28, 29]; one particularly interesting conclusion is that in the limit of the topic sequence, a single individual emerges holding all of the social power even in non-star graphs.

However, it has proved challenging to theoretically analyse the dynamics of (10) with the modification as in (18), if at least two or more individuals are somewhat stubborn, i.e.,  $\exists i, j : \lambda_i, \lambda_j < 1$ . For then, opinions do not reach a consensus for topic  $s$  even if  $\mathcal{G}[\mathcal{A}(s)]$  is strongly connected, as identified below Lemma 4. Thus, there is no convenient expression of the contribution of individual  $i$  using  $\zeta_i$  as in (9). One method is to use the matrix  $\mathbf{V}$ , see Lemma 4, which for a single topic is the mapping from the initial opinions to the final opinions. Thus, the equivalent of  $\zeta(s)^T$  in (9) is set to be

$$\zeta(s) = n^{-1} \mathbf{V}(s)^T \mathbf{1}_n \quad (19)$$

where  $\mathbf{V}(s) = (\mathbf{I}_n - \mathbf{A}(s))^{-1} (\mathbf{I}_n - \mathbf{A})$  and  $\mathbf{A}(s)$  is defined as in (8). To the authors' knowledge, the only paper treating this extension theoretically is the preliminary work in [66]. There, and assuming that  $\lambda_i < 1$ ,  $\forall i \in \{1, \dots, n\}$ , it was established that  $\mathbf{F}$  in (12) is replaced by  $\hat{\mathbf{F}} : \zeta(s) \rightarrow \zeta(s+1)$ , where  $\hat{\mathbf{F}}(\zeta)$  is a certain left eigenvector of  $\mathbf{U}$ , given by

$$\mathbf{U} = \frac{\mathbf{1}_n \mathbf{1}_n^T}{n} - (\mathbf{I}_n - \mathbf{A})^{-1} \mathbf{A} (\mathbf{I}_n - \text{diag}(\zeta_i)) (\mathbf{I}_n - \mathbf{C}). \quad (20)$$

In particular, the eigenvector  $\hat{\mathbf{F}}(\zeta)$  has all positive entries, satisfies  $\hat{\mathbf{F}}(\zeta)^T \mathbf{1}_n = 1$ , and is associated with the simple eigenvalue at  $\rho(\mathbf{U})$ . It should however be noted that  $\mathbf{U}$  is not necessarily a row-stochastic matrix, since there is no guarantee that the diagonal entries are non-negative, though it does have nonnegative off-diagonal entries. In fact,  $\mathbf{U}$  is a Metzler matrix, having certain special properties (see [47]).

## 4 Differences in expressed and private opinion

It has been well recognised in the literature from social psychology, sociology and the like that individuals interacting with one another may on occasions express opinions which are not consistent with their privately held opinions. This section is devoted to presenting a dynamic model which incorporates this distinction of opinions. This distinction often arises because an individual feels pressured in a group situation to conform to a social standard or norm. Studies identifying existence of this pressure, for the most part were initially qualitative in nature and go back at least 70 years. It was found that in

factories, group pressure can force individuals with high productivity rates to lower their rates to match a desired group standard<sup>[67]</sup>. Almost seventy years ago, Asch<sup>[40]</sup> conducted what is one of the most seminal experiments on social conformity, and this work is of great motivational value for our own work. Peer punishment is often threatened upon, or dealt to, individuals who deviate from behaviour acceptable to the group, e.g., in a gang<sup>[68]</sup>. This pressure can occur even when the norm is destructive to the group<sup>[69]</sup>.

Separately, the idea that an individual can simultaneously hold different private and public views on the same topic has been extensively documented. It was found in one extensive field study that over one third of jurors on criminal jury panels would have privately voted against the final decision of their jury panel<sup>[70]</sup>. Another recent work coined the term preference falsification to describe a situation where an individual knowingly, or subconsciously, expresses an altered form of his/her true opinion<sup>[41]</sup>. Unpopular norms can be enforced even if most individuals privately dislike them due to fears of isolation and exposure<sup>[50, 71]</sup>. The term pluralistic ignorance has been used to describe the outcome of the large scale occurrence of discrepancies between private and public opinions: individuals believe that the public majority support position *A* (because of their expressed opinions, as reported in the news media perhaps) when in reality the majority support (privately) position *B*, see [43, 72, 73].

Naturally, quantitative models have been proposed to try and capture some of the above phenomena. Not surprisingly, a number of models exist for Asch's experiments. These include the social influence model<sup>[74]</sup>, the norm influence model<sup>[75]</sup>, and the other-total-ratio model<sup>[76]</sup>; a number of these are summarised in the following meta-study<sup>[77]</sup>. However, these models are static, being essentially curve fitting for data from the Asch experiments. We are instead motivated to develop a dynamic agent-based model in line with those explored in the previous sections, as these can give us richer insight into how opinions can change over a network, including temporal change which has been noted as being extremely important<sup>[78]</sup>.

#### 4.1 EPO model

The model we present here is a novel one, inspired by Solomon Asch's seminal experiments on conformity under pressure<sup>[40]</sup>, and the many other prior works we discussed above. We term it the expressed-private-opinions (EPO) model. The model is dynamic (as opposed to the static ones discussed above) and seeks to study how differences between an individual's expressed and private opinions can arise<sup>[79]</sup>. In particular, we assume that an individual expresses an opinion which is the individual's private opinion altered due to a pressure to conform to the social network's average opinion. In other words, the

individual has some "resilience" to the pressure, but is not unaffected by it. We also assume each individual remains somewhat attached to the individual's initial opinion as in the Friedkin-Johnsen model. While the notion of "stubbornness" has been commonplace in opinion dynamics for some time, e.g., in the Friedkin-Johnsen model, the notion of "resilience" is introduced specifically for our model.

The mathematical form of the model can be seen as a modest adjustment of the Friedkin-Johnsen model. The key extension requires ascribing to the *i*-th agent a scalar parameter termed resilience, in the following way. The quantity  $x_i(k)$  constitutes the agent's private opinion at time *k*, and there is now a second quantity associated with the agent, namely, an expressed opinion  $\hat{x}_i(t)$  which other agents learn of, and it is derived by combining the private opinion and the effect of group pressure (effectively, the social network's average opinion, taking obviously the expressed rather than private opinions for the calculation of the average). This means that the update equations are as follows:

$$x_i(k+1) = \lambda_i[a_{ii}x_i(k) + \sum_{j \neq i}^n a_{ij}\hat{x}_j(k)] + (1-\lambda_i)x_i(0) \quad (21a)$$

$$\hat{x}_i(k) = \phi_i x_i(k) + (1-\phi_i)\hat{x}_{avg}(k-1) \quad (21b)$$

where  $a_{ij}$  and  $\lambda_i$  are the same as that defined in Section 2.2. Here  $x_{ave}(k) = \sum_{i=1}^n \frac{x_i(k)}{n}$  is called the public opinion as consistent with [80] and  $\phi_i \in [0, 1]$  is termed the resilience, i.e., the ability for individual *i* to withstand group pressure. It is instructive to observe that if  $\phi_i = 1$ , then the individual is fully resilient to pressure and the expressed and private opinions coincide, while if  $\phi_i = 0$ , the individual's own opinion is totally overwhelmed by the network's average opinion.

Notice that (21a) is almost identical in functional form to the Friedkin-Johnsen model, but only  $\hat{x}_j(k)$ , i.e., the neighbour's expressed opinions, and not  $x_j(k)$ , are available to individual *i*. This is a key departure from existing opinion dynamics models, which assume a single opinion per individual per topic (some models such as [31, 32, 81] assume simultaneous discussion of multiple topics, but an individual holds a single opinion per topic). From (21b), we observe that individual *i*'s expressed opinion  $\hat{x}_i(k)$  is a convex combination of his or her private opinion  $x_i(k)$  and the public opinion in the previous round of discussion,  $\hat{x}_{avg}(k-1)$ . The motivation is to capture the normative pressure to conform that arises in group situations<sup>[40, 42]</sup> that has been reported as driving an individual to express an opinion that is modified to be closer to the public average opinion. One can therefore consider  $(1-\phi_i)\hat{x}_{avg}(k-1)$  as a "force" that is exerted due to a pressure to conform.

**Remark 2.** It is perhaps remarkable that, despite the

many different advances in opinion dynamics modelling over the past several decades, almost no models exist which assume each individual has a private opinion  $x_i(k)$  and expressed opinion  $\hat{x}_i(k)$ . A model proposed in [50] does assume separate private and expressed opinions, but assumes that  $x_i, \hat{x}_i$  take on binary values; the model is more appropriate for modelling of unpopular norm enforcement as opposed to the evolution of opinions that take on values in a continuous interval. The influence, susceptibility, and conformity (ISC) model proposed in [23] does assume each individual has  $x_i, \hat{x}_i$  that take on values in a continuous interval, and is perhaps the closest in spirit to the one proposed in this paper. However, the updating mechanism for the opinions is fundamentally different, and the ISC model is extremely complex and nonlinear; it becomes incredibly difficult, if not impossible, to study the ISC model analytically. Moreover, we propose our model while aiming for a balance between simplicity for ease of analysis and complexity for capturing a wider range of phenomena than current models are able to. Last, a model with fewer parameters makes data fitting and parameter estimation in experimental validations a tractable objective; this is obvious from the fact that the Friedkin-Johnsen model is one of the few models with extensive empirical data<sup>[27–29]</sup>.

The reader may have noticed that each individual's expressed opinion depends on a global quantity  $\hat{x}_{avg}$ . For small-sized and medium-sized networks, it is not difficult to imagine that this is available. For large networks, such information might come from opinion polls, or trends in social media. However, it is also possible to consider the model with (21b) replaced by

$$\hat{x}_i(k) = \phi_i x_i(k) + (1 - \phi_i) \hat{x}_{i,avg}(k-1) \quad (22)$$

where  $\hat{x}_{i,avg} = \sum_{j \in \mathcal{N}_i} b_{ij} \hat{x}_j$  is the weighted average of the expressed opinions of individual  $i$ 's neighbours. Here,  $b_{ij} \geq 0$  are general weights satisfying  $\sum_{j=1}^n b_{ij} = 1$  and  $b_{ij} > 0 \Leftrightarrow a_{ij} > 0$ . (The nonzero  $b_{ij}$  corresponding to a fixed  $i$  may all be equal). We call  $\hat{x}_{i,avg}$  the local public opinion of individual  $i$ . Much of the analysis and results, detailed in the following subsection, of the model using the global public opinion, i.e., (21b), can be carried over to the local public opinion variant, i.e., (22); we refer the reader to [79] for details of the differences.

## 4.2 Analysis of the dynamics

Theoretical analysis of the model<sup>[79]</sup> has led to several conclusions, which are illustrated in an example simulation in Fig. 3. We now point out some of the most interesting results, detailing them in a more informal way in order to put the focus on interpretation of the results in a social context. Details of the exact analysis are found in [79]. The following conclusions were drawn under the assumption that  $\mathcal{G}[\mathbf{A}]$  is strongly connected, and that

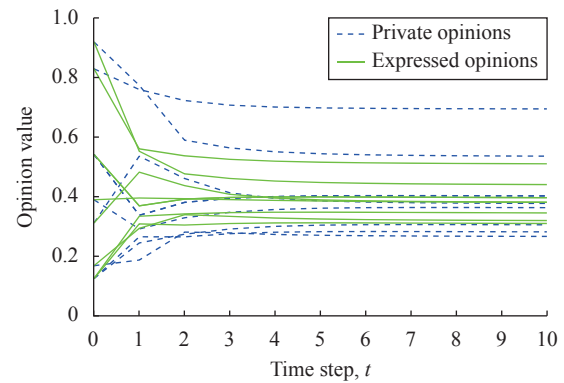


Fig. 3 Evolution of opinions for a network of 8 individuals. The opinions converge to a steady-state under mild assumptions on the network connectivity.

$\phi_i, \lambda_i \in (0, 1)$  for all  $i \in \{1, \dots, n\}$ . We are particularly interested in the evolution of the vector of expressed and private opinions of the individuals,  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_n]^T$  and  $\mathbf{x} = [x_1, \dots, x_n]^T$ , respectively.

1) The combination of a) pressure to conform to the public opinion  $\hat{x}_{ave}$ , b) stubborn attachment to the individual's initial opinion, and c) the strong connectedness of the network, means that a steady state is reached exponentially fast.

2) Interestingly, in general, any individual's private and expressed opinions are unequal at equilibrium, i.e.,  $x_i(\infty) \neq \hat{x}_i(\infty)$  for every  $i \in \{1, \dots, n\}$ .

3) Let us define disagreement in the private and expressed opinions at time  $k$  as  $V(k)$  and  $\hat{V}(k)$ , respectively, with

$$V(k) = \max_i x_i(k) - \min_i x_i(k)$$

$$\hat{V}(k) = \max_i \hat{x}_i(k) - \min_i \hat{x}_i(k).$$

It turns out that there is greater disagreement among the private opinions than expressed opinions at equilibrium, i.e.,  $V(\infty) > \hat{V}(\infty)$ . This is due to the effects of a pressure to conform to a social norm: people are more willing to voice agreement in a social network, but less willing to shift their private opinions. The smallest interval containing the private opinions actually “encloses” the smallest interval containing the expressed opinions at equilibrium.

4) It is possible to estimate a lower bound on the level of disagreement in the final private opinions,  $V(\infty)$  given the level of disagreement in the expressed opinions  $\hat{V}(\infty)$ , and an estimate of how resilient the individuals are to the pressure to conform. This is important, because large spreads of opinions in a group can be destructive of cohesiveness<sup>[82, 83]</sup>, and if known, can trigger some form of remedial action.

5) Though certainly not apparent in the figures, the steady state values of private and expressed opinions that are reached, while dependent on the initial private opin-

ions  $x(0)$ , are actually independent of the initial expressed opinions  $\hat{x}(0)$ . That is, no matter what individuals may say at the outset of a discussion, it is their initial private opinions only which determine both the steady state private and expressed opinions,  $x(\infty)$  and  $\hat{x}(\infty)$ .

6) The connectedness of the network leads to an individual's resilience to pressure to conform having a "propagating effect". Changing an individual's level of resilience to this pressure leads to a specific pattern of changes to every other individual's expressed opinion at equilibrium. More specifically, if the resilience  $\phi_j$  of individual  $j$  increases, so the individual becomes more reliant on his/her expressed opinion and less reliant on the group norm, the final expressed opinion  $\hat{x}_k(\infty)$  of any individual  $k$  with  $k \neq j$ , also becomes more influenced by the final private opinion  $x_j(\infty)$  of individual  $j$  than previously.

7) Whereas in the standard DeGroot model, the dynamical behaviour has the property that the maximum of all opinions is always monotone decreasing and the minimum of all such opinions is monotone increasing, no such property is guaranteed to hold for either expressed or private opinions in the EPO model, although in some cases the property may be present.

### 4.3 Asch's experiments and pluralistic ignorance

#### 4.3.1 Asch's experiments in the EPO model

The EPO model has also been used to accurately predict and explain Asch's conformity experiments<sup>[40]</sup>. Asch's conformity experiments aimed to capture how an individual reacts when he or she is faced with a unanimous group of other individuals who openly question his or her belief in an indisputable fact. The original experiment is illustrated in Fig. 4, where the individuals have to judge the lengths of lines, with 7 individuals (termed confederates in later literature) unanimously picking the wrong answer by private a priori arrangement with the experimenter. The reaction of the 8th, test, individual was recorded. Some test individuals would bend to the pressure of the unanimous group and also select the incorrect answer while others would resist; further, of those who bend, it may be that their private and expressed opinions bend in a dynamic process, or it may be that their private opinion remains little changed, whereas their expressed opinion changes substantially. We label without



Fig. 4 Example of the Asch experiment, with an illustrative influence network of eight individuals. The individuals are required to openly discuss the length of lines, and state their individual beliefs as to which one of  $A$ ,  $B$ ,  $C$  has the same length as the green line. Clearly  $A$  is equal in length to the green line. The confederates (seven blue individuals) unanimously express belief in the same wrong answer, e.g.,  $B$ .

loss of generality the test individual as individual 1 of the network, and show that each reaction of the test individual recorded by Asch can be accurately predicted by different values for the parameter pair  $\lambda_1, \phi_1$ , i.e., susceptibility to influence (or lack of stubbornness) and resilience to the group pressure (see [79] for details). Here, we summarise the outcomes in Table 1. In particular, we denote the confederate opinions as  $\hat{x}_c(k)$  and  $x_c(k)$ , and enforce that  $\hat{x}_c(k) = x_c(k)$  for all  $k \geq 0$ . This is because the confederates were told before the experiment to be stubborn and openly express belief in the wrong answer. The opinions of the test individual,  $x_1(k)$ ,  $\hat{x}_1(k)$  are set so that  $x_1(0) = \hat{x}_1(0)$  is the correct answer. This means that  $x_1(0) = \hat{x}_1(0) \neq \hat{x}_c(0) = x_c(0)$ , (and of course as time evolves the test individual's opinions in general move away from the initial values). Table 1 details how the opinions of the test individual evolve depending on his or her susceptibility and resilience parameters,  $\lambda_1$  and  $\phi_1$ .

#### 4.3.2 Pluralistic ignorance in the EPO model

Pluralistic ignorance, as introduced at the beginning of Section 4, is a social phenomenon where the majority of a population privately reject an opinion position, but people believe there is majority support for that position, e.g., in the 1960s, white Americans misidentified how much support there was for racial segregation<sup>[72]</sup>. In Fig. 5, our model shows that stubborn extremists (zealots) placed at well connected nodes in a scale-free network can create massive pluralistic ignorance in the general population. Not only does this create confusion and misinformation about the true desires of the population, it is known that large differences in expressed and private opinions that are sustained for a long time can foster discontent among individuals, leading to unexpected and drastic actions<sup>[82, 83]</sup>.

Table 1 Types of test individuals, their associated parameters and the individual's opinion change

	$\lambda_1$	$\phi_1$	Change in opinions
Independent	Low	High	$x_1(\infty) \approx \hat{x}_1(\infty) \approx x_1(0)$
Yielding, distortion of judgment	High	Any value in $(0, 1)$	$x_1(\infty) \approx \hat{x}_1(\infty) \approx \hat{x}_c(0)$
Yielding, distortion of action	Low	Low	$x_1(\infty) \approx x_1(0)$ , but $\hat{x}_1(\infty) \approx \hat{x}_c(0)$



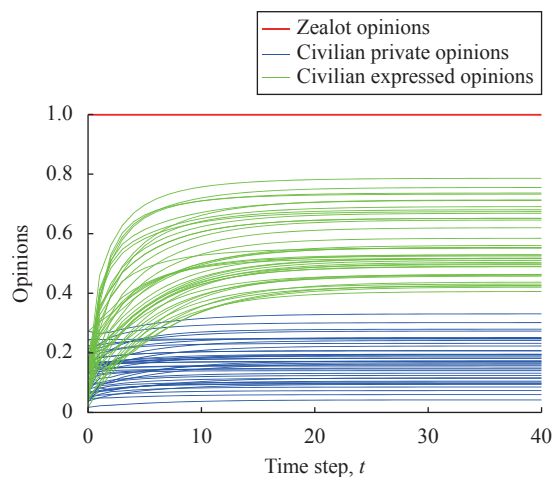


Fig. 5 Temporal evolution of opinions for a network of 200 individuals. There are 5 zealots, and for clarity, only 40 civilian (non-zealot) individuals are shown. Stubbornness of the civilians leads to only minor changes in their private opinions, but the zealots spread misinformation, causing the civilians' expressed opinions to shift significantly.

We have also verified that our model can capture the pluralistic ignorance phenomenon observed in field studies by Prentice and Miller<sup>[43]</sup>, of attitudes to consumption of alcohol on Princeton University campus. In [43], the authors surveyed a large cohort of undergraduate students, asking each of them their opinions on the alcohol drinking culture on campus. Several notable aspects were recorded, each of which appears when we place the study in the framework of our model.

Prentice and Miller<sup>[43]</sup> found that men and women responded differently to sustained discrepancies in private and expressed opinions over time (the presence of time evolution was heavily stressed in the study, and helps support the relevance of our model over existing static models for conformity). Women continued to display pluralistic ignorance over time. At the start of the survey, women generally were uncomfortable with the drinking culture (their private opinions), but assumed everyone else was comfortable with it (the expressed opinions of others). After several months, this had not changed by much. On the other hand, men at the beginning of the survey also returned similar opinion profiles as women, but after several months, had become affected by the pluralistic ignorance. In particular, after several months, men reported that they were comfortable with the drinking culture (their private opinions), and also assumed everyone else was comfortable with it (the expressed opinions of others). In other words, the private opinions of most male students changed over time due to the pluralistic ignorance, whereas those of female students did not. In our model, both male and female students had low resilience  $\phi_i$ , while male and female students had high and low susceptibility  $\lambda_i$ , respectively. Both types of students could co-exist in the same network, as was observed in the field experiment.

By way of general comment, we note that the influence of zealots can be substantially affected by several network properties: the first is the topology, such as whether the network is a small-world network, or a scale-free network. It turns out that in a small world network, zealots do not have to reach far to affect a nonzealot, although the zealot's influence may be diluted because nonzealots have a number of other nonzealot neighbours. In a scale-free network, there tend to be nodes with very high degree, and if, and perhaps only if, zealots can be deliberately positioned at such a node, they can be very powerful. Whether average public opinions seen by an agent are local or global is also relevant, i.e.,  $\hat{x}_{i,lavg}$  or  $\hat{x}_{avg}$  as in (22) or (21b). If an agent seeks just the average opinion of his/her neighbours, or one and two-hop neighbours, he may see no average opinion affected by a zealot. The influence of zealots will take time to spread out, and perhaps become attenuated in the process.

#### 4.4 Future work on the EPO model

A number of exciting future directions exist for the EPO model, which is still very new. Aside from the usual extensions to consider time-varying interactions, gossip-based algorithms and incorporation of other accepted features such as bounded-confidence or antagonistic interactions, we briefly cover two specific extensions to interesting social phenomena.

Phenomena and behaviour related to pluralistic ignorance are of particular interest in studying how misinformation spreads through high profile media figures, or hostile bot accounts on social media (such as the Twitter bots of a foreign country commenting on political matters). We also expect this to be closely linked with the concept of the spiral of silence<sup>[80, 84]</sup>, which is a social phenomenon where an individual is more likely to stop expressing his or her opinion if that individual perceives (rightly or wrongly) that everyone else is moving away from his or her opinion. We hope that study and development of these models (including incorporation of a spiral of silence mechanism, perhaps by inclusion of an event-based communication component) will reveal to us deeper insight about the role extremists play in creating a divergence in the private and expressed opinions of the general population, and provide some guidance for developing effective countermeasures. For example, it is not immediately clear whether it is better to 1) introduce a new set of extremists that are on the opposite side of the opinion spectrum to the original zealots (which might risk polarising the network), or 2) spread more moderate opinions (which might risk being lost in the presence of the original zealots).

It would also be of great interest to study such models for online social networks, where the 1-9-90 rule is prevalent in many discussion forums. Roughly speaking, 1% of users create content, 9% of users participate by post-

ing opinions, and 90% of users do not participate at all<sup>[85]</sup>; how do the private opinions of the 90% evolve as a result of the 10% expressing opinions, and what happens if these ratios change?

## 5 Opinion dynamics on logically related issues

What do we mean by the term “logically related issues”? We provide an example. Suppose that a topic is the truthfulness of a statement, and a positive (respectively negative) opinion value indicates belief that the statement is true (respectively false), with the magnitude of the opinion value indicating strength of conviction. For this section, we assume the opinions lie in  $[-1, 1]$ . Consider two topics being simultaneously discussed; 1) mentally challenging tasks are just as exhausting as physically challenging tasks and 2) that Go and chess should be considered sports in the Olympics. Clearly a person who believes topic 1 is true is more likely to believe topic 2 is true. Let individual  $i$ 's opinion vector be  $\mathbf{x}_i = [x_i^1, x_i^2]^T$ . (We use the bold face to distinguish this situation from the earlier problems which restrict consideration at any one time to a single topic.) For topic 1, if  $x_i^1$  is positive (respectively negative), then individual  $i$  believes mentally challenging tasks are just as exhausting (respectively not as exhausting) as physically challenging tasks. For topic 2, if  $x_i^2$  is positive (respectively negative) then individual  $i$  believes (to some degree) Go and chess should be considered (respectively not considered) an Olympic sports. To capture in a model how individuals interact when interdependent topics are involved, Parsegov et al.<sup>[31, 32]</sup> proposed to combine the idea of the DeGroot model with the notion of a logic matrix, which captures logical interdependence of the topics, and which for individual  $i$  is denoted as  $\mathbf{C}_i$ . With  $\mathbf{A} = (a_{ij})$  a stochastic matrix capturing the network effect of interaction among individuals as in the DeGroot model (see Section 2.2), one has

$$\mathbf{x}_i(k+1) = \sum_{j=1}^n a_{ij} \mathbf{C}_i \mathbf{x}_j(k). \quad (23)$$

To clarify the purpose of the logic matrix, consider temporarily an individual  $i$  with no neighbours, so that (23) becomes

$$\mathbf{x}_i(k+1) = \mathbf{C}_i \mathbf{x}_i(k) \quad (24)$$

and consider the above mentioned example regarding Go and chess as Olympic sports. An individual  $i$  may hold the belief that an event can only be an Olympic sport if it is exhausting. One possible logic matrix is given by

$$\mathbf{C}_i = \begin{bmatrix} 1 & 0 \\ 0.7 & 0.3 \end{bmatrix}. \quad (25)$$

Suppose that  $\mathbf{x}_i(0) = [1, -0.7]^T$ , i.e., initially agent  $i$  believes mentally challenging tasks are as exhausting as physically challenging tasks, but does not believe Go and chess should be Olympic sports. According to (24),  $x_i^1(1) = x_i^1(0)$  and  $x_i^2(1) = 0.7x_i^1(0) + 0.3x_i^2(0) = 0.49$ . That is agent  $i$ 's evaluation on the logical interdependence of topic 2 on topic 1 (because agent  $i$  holds the view that an event can only be an Olympic sport if it is exhausting) causes  $x_i^2(k)$  to shift from  $-0.7$  to  $0.49$ ; thus, agent  $i$ 's opinion on topic 2 in one step has been altered due to  $i$ 's opinion on topic 1. Eventually,  $\lim_{k \rightarrow \infty} \mathbf{x}_i(k) = [1, 1]^T$ , and agent  $i$  comes to believe (after some internal reflection) that Go and chess should be Olympic sports. On the other hand, if agent  $i$  does not initially believe mentally challenging tasks are as exhausting as physically challenging tasks, e.g.,  $x_i^1(0) = -0.8$ , then  $\lim_{k \rightarrow \infty} \mathbf{x}_i(k) = [-0.8, -0.8]^T$  and agent  $i$  believes chess should not be an Olympic sport.

Another individual  $j$  might take the view that Go and chess being Olympic sports implies that mentally challenging tasks are to some degree as exhausting as physically challenging tasks and so may have

$$\mathbf{C}_j = \begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix}. \quad (26)$$

The nonzero 12 entry indicates that agent  $j$  views topic 1 as being logically dependent on topic 2.

While the above  $\mathbf{C}_i$  and  $\mathbf{C}_j$  are row-stochastic, we do not in general require  $\mathbf{C}_i$  to be row-stochastic (though other constraints will apply, depending on whether we are working with continuous-time or discrete-time models). For convenience, we will restrict attention to one possibility, namely discrete-time models. What properties then should a matrix  $\mathbf{C}_i$  have? It is crucial that in the absence of inputs from other individuals, i.e., with dynamics (24), an individual's own belief system should be consistent, in the sense of giving rise to a steady-state opinion vector. Thus, (24) for arbitrary initial conditions with each entry confined to  $[-1, 1]$  should have a transient solution, including a limiting solution, such that the transient and limit also have entries confined to this interval. The requisite conditions are set out in detail in [31, 86], and we report them here in the following assumption. (A parallel restriction applies in continuous time of course<sup>[81]</sup>.)

**Assumption 2.** The matrix  $\mathbf{C}_i$ , for each  $i$ , has either

- 1) A semi-simple<sup>4</sup> eigenvalue 1. All other eigenvalues of  $\mathbf{C}_i$  have modulus strictly less than 1, or
- 2) All eigenvalues of  $\mathbf{C}_i$  have modulus less than 1.

Moreover, for all  $i \in \mathcal{I}$  and  $p \in \mathcal{J}$ ,  $\sum_{q=1}^m |c_{pq,i}| = 1$ . If topic  $p$  is independent of all other topics, i.e.,  $c_{pq,i} = 0$  for all  $q \neq p$ , then  $c_{pp,i} = 1$ . The diagonal entries are nonneg-

<sup>4</sup>By semi-simple, we mean that the geometric and algebraic multiplicities are the same. Equivalently, the Jordan blocks of the eigenvalue 1 are all 1 by 1.

ative,  $c_{pp,i} \geq 0$ .

## 5.1 Disagreement due to logical differences

Unsurprisingly, if all  $C_i$  are the same, from an algebraic point of view, equation (23) for the entire network is rather easy to analyse, with the aid of Kronecker products. Indeed, Parsegov et al.<sup>[31]</sup> considers networks of individuals all having the same  $C_i$  (along with other aspects including stubbornness as described by the Friedkin-Johnsen model, see (3)). The work<sup>[32]</sup> does consider heterogeneous  $C_i$ , but focuses on obtaining convergence rather than investigating the effects of heterogeneity. Moreover, the stability result in [32] requires at least one individual to have some stubbornness. Our work<sup>[86]</sup> has sought to focus on what happens to the final opinion distribution when the  $C_i$  are distinct, and thus makes the assumption that the individuals are not stubborn; we now record several key outcomes which are different from those in [31, 32].

Notice that the matrix  $C_i$  in (25) is lower triangular. This is common in the social context of the situation, reflecting the fact that individual  $i$ 's belief system flows from one or more axioms, or truths which they consider to be indisputable. Our work begins by focusing on such lower triangular  $C_i$ .

First, we establish the possibly nonobvious fact that all models of this type are convergent, given the standard connectivity condition in opinion dynamics models of having  $\mathcal{G}[\mathbf{A}]$  strongly connected and aperiodic, and Assumption 2 imposed on the logic matrix  $C_i$ , arising from the need for (24) to describe a consistent belief structure.

Next, we consider the question of whether consensus on individual topics can be achieved. Suppose to begin that there are two topics under discussion, and that topic 2 depends on topic 1 but topic 1 does not depend on topic 2, for every individual. Thus,

$$C_i = \begin{bmatrix} 1 & 0 \\ c_{21,i} & c_{22,i} \end{bmatrix}. \quad (27)$$

The major result for this case is as follows. A proof of this result can be found in [86].

**Theorem 2.** Suppose the graph  $\mathcal{G}(\mathbf{A})$  is strongly connected and that individuals are discussing two logically interdependent topics according to (23). Suppose further that the logic matrices of all individuals are of the form of (27), with Assumption 2 holding. Then:

- 1) Consensus of topic 1 always occurs;
- 2) For almost all initial conditions, consensus of topic 2 occurs if and only if no two 21 entries of the set of  $C_i$  have different signs.

Evidently, if all individuals have the same sign for the 21 entry of their  $C_i$  matrix, i.e., there are no competing logical interdependence structures, then the dynamic changes in each individual's collection of opinions are pro-

duced by consistent action of the logic matrix (the individual's belief system) and the network (the other individuals). As a result, consensus is reached in steady state on both topics. On the other hand, if there are two individuals  $p, q$  for which  $c_{21,p}$  and  $c_{21,q}$  have opposite signs, then consensus will not be achieved on topic 2. Evidently, major but not complete differences in belief structures will not destroy consensus.

This is illustrated in Figs.6 and 7, where there are 6 individuals discussing 2 topics, with  $C_i$  taking the form of (27). In Fig.6, individuals 1, 2, 3 have the same  $C_i = \hat{C}$  matrix, while individuals 4, 5, 6 have the same  $C_j = \tilde{C}$  matrix. The two logic matrices satisfy  $\hat{C} \neq \tilde{C}$  but have the same sign pattern;  $c_{21,i}$  is negative and  $c_{22,i}$  is positive for all  $i$ . Fig.6 shows that the individuals reach a consensus for each of topics 1 and 2, with the consensus values for the two topics being different. Fig.7 shows a simulation where five individuals  $i = 1, 2, \dots, 5$  have the same  $C_i$ , and whose opinions on topic 2 are represented by the solid blue lines. The sixth individual  $j = 6$ , whose opinion on topic 2 is indicated by the dashed blue line, has  $C_j$  with  $c_{21,j}$  having opposite sign to  $c_{21,i}$ . As can be seen, opinions converge to a persistent disagreement, even for the five individuals with the same logic matrix  $C_i$ .

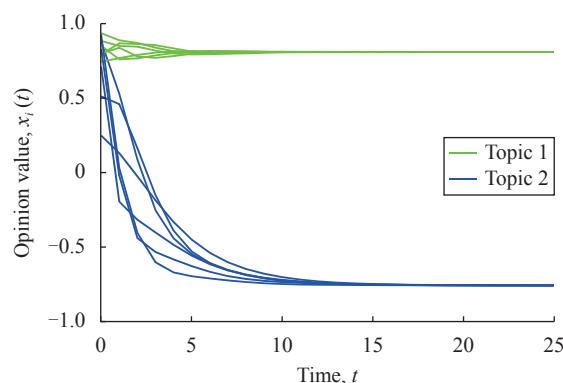


Fig.6 Evolution of 6 individuals' opinions for two logically interdependent topics, with topic 1 independent, and topic 2 dependent on topic 1. The individuals have heterogeneous  $C_i$ , but all  $C_i$  matrices have the same sign pattern, i.e., there are no competing logical interdependencies.

The situation with three or more topics is more subtle. While convergence can be established much as for the case of two topics, the question of consensus is only partially resolved, see [86]. Suppose that all  $C_i$  have the form

$$C_i = \begin{bmatrix} 1 & 0 & 0 \\ c_{21,i} & c_{22,i} & 0 \\ c_{31,i} & c_{32,i} & c_{33,i} \end{bmatrix}. \quad (28)$$

Then discrepancies in signs in the second row between corresponding elements will in general prevent consensus from being reached on topic 3 as well as on topic 2, suggesting that the cascade structure has a prominent im-

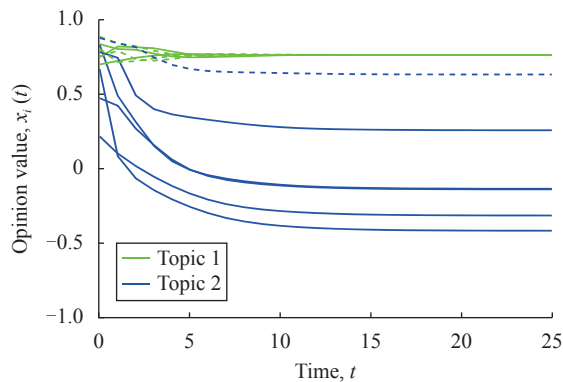


Fig. 7 Evolution of 6 individuals' opinions for two logically interdependent topics, with topic 1 independent, and topic 2 dependent on topic 1. The solid blue lines represent individuals with the same  $C_i$  matrix, and the dotted blue line represents an individual  $j$  whose 21 entry of  $C_j$  has opposite sign to the 21 entry of the other  $C_i$ .

pact. Also minor differences between individuals of value (even without sign difference) in any of the entries of the third row can lead to lack of consensus in topic 3. Thus there is a significant distinction to be drawn between the two topic case and the three (or more) topic case.

The failure of consensus is not just minor; there can arise a diverse disagreement of opinions from minor differences in the values of the entries. In other words, when consensus fails to be achieved for a given topic using the model in (23), the network tends to settle into a state of strong diversity. Recall from the Introduction that on strongly connected graphs, some models give rise to weak diversity (e.g., Hegselmann-Krause, and Altafini) while other models give rise to strong diversity (e.g., Friedkin-Johnsen), and it is strong diversity which is of greater interest. Our reported findings are of particular interest because strong diversity in the final opinions, e.g., in Fig. 7, can apparently be attributed to differences in logical interdependence between individuals' belief systems. To elaborate, strong diversity occurs even though each individual uses (23) in an attempt to bring their opinions closer together, and there is no presence of stubbornness<sup>[26]</sup>, bounded-confidence<sup>[10]</sup>, negative influence<sup>[16]</sup>, or the other usual processes to which disagreement may be attributed.

## 5.2 Results for continuous-time model

As a general rule, much the same set of conclusions arise in social network modelling whether one uses a continuous-time or discrete-time model. In the case of multiple interdependent topics however, we need to note one clear difference as identified in [81]. When one moves to continuous time, one can consider a time scale associated with the internal cognitive process for an individual to obtain a consistent belief system (i.e., the continuous time version of (24)), and one can consider a time scale

associated with the individual's interactions with other individuals, thereby causing some modification, and with the aim of achieving consensus between individuals. In the event that the second time scale dominates the first, we have been able to show theoretically, and validated the conclusion with simulations, that instability may result. Put another way, one must not rush the process of arriving at consensus between individuals when there are multiple related topics to consider. People can only adjust to the opinions of others so fast while also trying to secure logical consistency: if one tries to respond to others' opinions faster, fracture and instability may result – at least in the mathematical model. This phenomenon may be related to cognitive overload, in which an individual's internal capacity for processing external information can become overwhelmed if that information is received at a high density. As a result, the individual suffers a major decrease in decision-making and cognitive abilities<sup>[87, 88]</sup>.

The possibility of instability just identified cannot be duplicated in a discrete-time model. While perhaps paradoxical, there is an explanation. Because the fact a sampling time exists in a discrete-time model sets an upper limit on the speed with which change can be achieved, and the rate at which interactions with other individuals affects an individual's opinion cannot become an arbitrarily large multiple of the rate at which the individual's belief system processes information.

## 5.3 Future work on the opinion dynamics of logically related issues

It will be evident from the preceding material that there are significant gaps in our understanding of the opinion dynamics when individuals are discussing logically related issues. One flagged above is to provide a comprehensive theory when there are three or more topics in any study of multiple related topics. Furthermore, we hope to explore other logic matrix structures. For example, irreducible logic matrices appear to not have the same tendency to generate disagreement even when there is heterogeneity. This may provide insight into which belief structures are more robust in generating consensus, and which belief structures are more likely to lead to disagreement. Eventually, one would like to have a comprehensive account whereby, given a certain logical interdependence structure, we are able to precisely quantify the impacts of the logical structure on the final opinion distribution and identify which topics are likely to reach a consensus, and which topics will show a strong diversity of opinions.

In a separate direction, it would be of interest to continue investigations on the continuous-time counterpart model. A comprehensive convergence result for networks where individuals have some stubbornness is still lacking. The interplay between stubbornness and cognitive over-



load is also of interest; simulations and preliminary results have suggested that high levels of stubbornness can stabilise an opinion dynamical system in which absence of stubbornness would result in instability. Moreover, the same questions on the effects of heterogeneity in the logic matrices considered in the discrete-time model are equally worthy of study for the continuous-time model.

## 6 Conclusions

The oldest dynamical models for the evolution of opinions among a group of individuals go back some 60 years. On the time scales associated with scientific research, this corresponds to many generations, but some might say progress has not been fast, at least till recent times. More recent progress has been assisted by applications drivers. The desire to understand matters such as how customers or employees are motivated, or how bots or zealots may seek to disrupt an electoral process, are obvious examples which have stimulated many social scientists, including of course those working quite independently of engineers. At the same time recent developments in systems and control oriented towards studying networks, often in the first instance networks of physical entities such as robots, unmanned airborne vehicles and the like, have been another contributor stimulating great interest and at times rapid development in opinion dynamics, at least in the control community.

Many, but of course not all, developments in opinion dynamics need to be grounded in rigorous mathematics, as arises in fields such as control theory, especially networked control theory. Developments also need to make contact with the experimental side of opinion dynamics, for motivation and for validation. The raw material for much of this already exists in the literature, where one can find descriptions of laboratory experiments or field studies. Much of this literature however studies a static situation (perhaps a steady state situation at the end of a dynamic evolution), in contrast perhaps to the predisposition of a control scientist or engineer interested in time evolution of a system or its dynamic behaviour as well as its long term steady state.

Control engineering also has a long experience of seeking parsimonious models of physical systems, i.e., models with often fewer parameters than might be necessary to cover every single aspect of the system, but models which at the same time are useful for explanatory and predictive purposes. A focus on keeping models simple in the sense of minimising the number of parameters is likely to be important in future developments of opinion dynamics.

In this paper, we have covered three different extensions of the basic DeGroot model. In relation to the first, the evolution of social power, we stress that there are experimental studies (not conducted by us, but referenced in this paper) which validate the model. It passes the simplicity test. And we have indicated how, with addi-

tional modest parameters, behavioural characteristics can be used to build a more sophisticated model, although we lack experimental validation of this extended model. The second extension of the basic DeGroot and Friedkin-Johnsen models, dealing with differing expressed and private opinions, is a dynamic model that again passes the simplicity test, in some contrast perhaps to models in the social science literature of greater complexity and which pay less attention to conclusions potentially obtainable by system theoretic tools. The ability of this model to capture much that has been previously reported in the literature is striking. Then, we have examined a situation frequently encountered in real life, but not greatly examined yet in terms of opinion dynamics, i.e., the evolution of multiple opinions on logically related topics within a group of individuals. This work is both less theoretically developed and less supported currently by laboratory or field data.

It is appropriate to recall the caution that for large-scale networks, the DeGroot model and models derived from it may not be as effective as kinetic models. Self-evidently, the weighting parameters in the influence matrix will never be all obtainable, and for networks of even 1000 agents, the notion of working with a million parameters in the influence matrix is challenging, to say the least. One must therefore be aware of the scenarios for which agent-based models are, and are not, appropriate for modelling of opinion dynamics. Leaving aside the issue of network size, we hope we have made the case for the relevance of the work, for the power of the dynamic models, and for the sheer volume of scientific challenges still remaining to be addressed, even when in this paper we have focused on just a few of the recent advances.

## Acknowledgements

We would like to thank all of our co-authors of related work whose ideas contributed immeasurably to the themes covered in this paper: Chang-Bin Yu, Ji Liu, Tamer Başar, Ming Cao, Hyo-Sung Ahn, Yu-Zhen Qin, Alain Govaert, Minh Hoang Trinh and Young-Hun Lim.

This work was supported by the Australian Research Council (ARC) (No. DP-160104500) and Data61-CSIRO, Australia. This work was also supported in part by the European Research Council (No. ERC-CoG-771687) and the Netherlands Organization for Scientific Research (No. NWO-vidi-14134).

## Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- [1] J. R. P. French Jr. A formal theory of social power. *Psychological Review*, vol.63, no.3, pp.181–194, 1956. DOI: 10.1037/h0046123.
- [2] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, vol.69, no.345, pp.118–121, 1974. DOI: 10.1080/01621459.1974.10480137.
- [3] J. Becker, D. Brackbill, D. Centola. Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences of the United States of America*, vol.114, no.26, pp.E5070–E5076, 2017. DOI: 10.1073/pnas.1615978114.
- [4] A. G. Chandrasekhar, H. Larreguy, J. P. Xandri. Testing models of social learning on networks: Evidence from a framed field experiment, Technical Report, Massachusetts Institute of Technology, Cambridge, USA, 2012.
- [5] G. Toscani. Kinetic models of opinion formation. *Communications in Mathematical Sciences*, vol.4, no.3, pp.481–496, 2006. DOI: 10.4310/CMS.2006.v4.n3.a1.
- [6] B. Düring, M. T. Wolfram. Opinion dynamics: Inhomogeneous Boltzmann-type equations modelling opinion leadership and political segregation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol.471, no.2182, Article number 20150345, 2015. DOI: 10.1098/rspa.2015.0345.
- [7] G. Albi, L. Pareschi, M. Zanella. Opinion dynamics over complex networks: Kinetic modelling and numerical methods. *Kinetic and Related Models*, vol.10, no.1, pp.1–32, 2017. DOI: 10.3934/krm.2017001.
- [8] G. Toscani, A. Tosin, M. Zanella. Opinion modeling on social media and marketing aspects. *Physical Review E*, vol.98, no.2, Article number 022315, 2018. DOI: 10.1103/PhysRevE.98.022315.
- [9] B. Chang, T. Xu, Q. Liu, E. H. Chen. Study on information diffusion analysis in social networks and its applications. *International Journal of Automation and Computing*, vol.15, no.4, pp.377–401, 2018. DOI: 10.1007/s11633-018-1124-0.
- [10] R. Hegselmann, U. Krause. Opinion dynamics and bounded confidence models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, vol.5, no.3, Article number 2, 2002.
- [11] V. D. Blondel, J. M. Hendrickx, J. N. Tsitsiklis. On Krause's multi-agent consensus model with state-dependent connectivity. *IEEE Transactions on Automatic Control*, vol.54, no.11, pp.2586–2597, 2009. DOI: 10.1109/TAC.2009.2031211.
- [12] J. Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, vol.18, no.12, pp.1819–1838, 2007. DOI: 10.1142/S0129183107011789.
- [13] W. Su, G. Chen, Y. G. Hong. Noise leads to quasi-consensus of Hegselmann-Krause opinion dynamics. *Automatica*, vol.85, pp.448–454, 2017. DOI: 10.1016/j.automatica.2017.08.008.
- [14] S. R. Etesami, T. Başar. Game-theoretic analysis of the Hegselmann-Krause model for opinion dynamics in finite dimensions. *IEEE Transactions on Automatic Control*, vol.60, no.7, pp.1886–1897, 2015. DOI: 10.1109/TAC.2015.2394954.
- [15] A. Mirtabatabaei, F. Bullo. Opinion dynamics in heterogeneous networks: Convergence conjectures and theorems. *SIAM Journal on Control and Optimization*, vol.50, no.5, pp.2763–2785, 2012. DOI: 10.1137/11082751X.
- [16] C. Altafini. Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, vol.58, no.4, pp.935–946, 2013. DOI: 10.1109/TAC.2012.2224251.
- [17] A. V. Proskurnikov, A. S. Matveev, M. Cao. Opinion dynamics in social networks with hostile camps: Consensus vs. polarization. *IEEE Transactions on Automatic Control*, vol.61, no.6, pp.1524–1536, 2016. DOI: 10.1109/TAC.2015.2471655.
- [18] J. Liu, X. D. Chen, T. Başar, M. A. Belabbas. Exponential convergence of the discrete-time and continuous-time Altafini models. *IEEE Transactions on Automatic Control*, vol.62, no.12, pp.6168–6182, 2017. DOI: 10.1109/TAC.2017.2700523.
- [19] W. G. Xia, M. Cao, K. H. Johansson. Structural balance and opinion separation in trust-mistrust social networks. *IEEE Transactions on Control of Network Systems*, vol.3, no.1, pp.46–56, 2016. DOI: 10.1109/TCNS.2015.2437528.
- [20] D. Cartwright, F. Harary. Structural balance: A generalization of Heider's theory. *Psychological Review*, vol.63, no.5, pp.277–293, 1956. DOI: 10.1037/h0046049.
- [21] M. Mäs, A. Flache, J. A. Kitts. Cultural integration and differentiation in groups and organizations. *Perspectives on Culture and Agent-based Simulations: Integrating Cultures*, V. Dignum, F. Dignum, Eds., Cham, Germany: Springer, pp.71–90, 2014. DOI: 10.1007/978-3-319-01952-9\_5.
- [22] G. D. Shi, A. Proutiere, M. Johansson, J. S. Baras, K. H. Johansson. The evolution of beliefs over signed social networks. *Operations Research*, vol.64, no.3, pp.585–604, 2016. DOI: 10.1287/opre.2015.1448.
- [23] P. Duggins. A Psychologically-motivated model of opinion change with applications to American politics. *Journal of Artificial Societies and Social Simulation*, vol.20, no.1, Article number 13, 2017. DOI: 10.18564/jasss.3316.
- [24] P. Dandekar, A. Goel, D. T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences of the United States of America*, vol.110, no.15, pp.5791–5796, 2013. DOI: 10.1073/pnas.1217220110.
- [25] V. Amelkin, F. Bullo, A. K. Singh. Polar opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, vol.62, no.11, pp.5650–5665, 2017. DOI: 10.1109/

- TAC.2017.2694341.
- [26] N. E. Friedkin, E. C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, vol. 15, no. 3–4, pp. 193–206, 1990. DOI: 10.1080/0022250X.1990.9990069.
  - [27] N. E. Friedkin, E. C. Johnsen. *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*, New York, USA: Cambridge University Press, 2011.
  - [28] N. E. Friedkin, P. Jia, F. Bullo. A theory of the evolution of social power: Natural trajectories of interpersonal influence systems along issue sequences. *Sociological Science*, vol. 3, pp. 444–472, 2016. DOI: 10.15195/v3.a20.
  - [29] N. E. Friedkin, F. Bullo. How truth wins in opinion dynamics along issue sequences. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 43, pp. 11380–11385, 2017. DOI: 10.1073/pnas.1710603114.
  - [30] C. C. Childress, N. E. Friedkin. Cultural reception and production: The social construction of meaning in book clubs. *American Sociological Review*, vol. 77, no. 1, pp. 45–68, 2012. DOI: 10.1177/0003122411428153.
  - [31] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, N. E. Friedkin. Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2270–2285, 2017. DOI: 10.1109/TAC.2016.2613905.
  - [32] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, S. E. Parsegov. Network science on belief system dynamics under logic constraints. *Science*, vol. 354, no. 6310, pp. 321–326, 2016. DOI: 10.1126/science.aag2624.
  - [33] A. V. Proskurnikov, R. Tempo. A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017. DOI: 10.1016/j.arcontrol.2017.03.002.
  - [34] A. V. Proskurnikov, R. Tempo. A tutorial on modeling and analysis of dynamic social networks. Part II. *Annual Reviews in Control*, vol. 45, pp. 166–190, 2018. DOI: 10.1016/j.arcontrol.2018.03.005.
  - [35] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, J. Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 4, Article number 2, 2017. DOI: 10.18564/jasss.3521.
  - [36] P. Jia, A. MirTabatabaei, N. E. Friedkin, F. Bullo. Opinion dynamics and the evolution of social power in influence networks. *SIAM Review*, vol. 57, no. 3, pp. 367–397, 2015. DOI: 10.1137/130913250.
  - [37] M. Ye, J. Liu, B. D. O. Anderson, C. B. Yu, T. Başar. On the analysis of the DeGroot-Friedkin model with dynamic relative interaction matrices. *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 11902–11907, 2017. DOI: 10.1016/j.ifacol.2017.08.1426.
  - [38] M. Ye, J. Liu, B. D. O. Anderson, C. B. Yu, T. Başar. Evolution of social power in social networks with dynamic topology. *IEEE Transactions on Automatic Control*, vol. 63, no. 11, pp. 3793–3808, 2019. DOI: 10.1109/TAC.2018.2805261.
  - [39] B. D. O. Anderson, M. Ye. Nonlinear mapping convergence and application to social networks. In *Proceedings of European Control Conference*, Limassol, Cyprus, pp. 557–562, 2018.
  - [40] S. E. Asch. Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership and Men*, H. Guetzkow, Ed., Oxford, UK: Carnegie Press, pp. 222–236, 1951.
  - [41] T. Kuran. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*, Cambridge, USA: Harvard University Press, 1997.
  - [42] R. L. Gorden. Interaction between attitude and the definition of the situation in the expression of opinion. *American Sociological Review*, vol. 17, no. 1, pp. 50–58, 1952. DOI: 10.2307/2088359.
  - [43] D. A. Prentice, D. T. Miller. Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology*, vol. 64, no. 2, pp. 243–256, 1993. DOI: 10.1037/0022-3514.64.2.243.
  - [44] P. E. Converse. The nature of belief systems in mass publics. *Ideology and Discontent*, D. Apter, Ed., New York, USA: Free Press, pp. 206–261, 1964.
  - [45] F. Bullo, J. Cortés, S. Martínez, *Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms*, Princeton, USA: Princeton University Press, 2009.
  - [46] C. D. Godsil, G. Royle, *Algebraic Graph Theory*, New York, USA: Springer, 2001.
  - [47] A. Berman, R. J. Plemmons. Nonnegative matrices. *Nonnegative Matrices in the Mathematical Sciences. Classics in Applied Mathematics*, Philadelphia, USA: SIAM, 1979.
  - [48] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation*, vol. 1, no. 4, Article number 19, 2013. DOI: 10.1145/2538508.
  - [49] A. Nowak, J. Szamrej, B. Latané. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, vol. 97, no. 3, pp. 362–376, 1990. DOI: 10.1037/0033-295X.97.3.362.
  - [50] D. Centola, R. Willer, M. Macy. The Emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, vol. 110, no. 4, pp. 1009–1040, 2005. DOI: 10.1086/427321.
  - [51] J. M. Hendrickx, G. D. Shi, K. H. Johansson. Finite-time consensus using stochastic matrices with positive diagonals. *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1070–1073, 2015. DOI: 10.1109/TAC.2014.2352691.
  - [52] A. Nedić, J. Liu. On convergence rate of weighted-averaging dynamics for consensus problems. *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 766–781, 2017. DOI: 10.1109/TAC.2016.2572004.
  - [53] W. Ren, R. W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005. DOI: 10.1109/TAC.2005.846556.
  - [54] M. Cao, A. S. Morse, B. D. O. Anderson. Reaching a consensus in a dynamically changing environment: A graphical approach. *SIAM Journal on Control and Optimization*, vol. 47, no. 2, pp. 575–600, 2008. DOI: 10.1137/060657005.
  - [55] N. E. Friedkin, E. C. Johnsen. Social influence networks

- and opinion change. *Advances in Group Processes*, vol. 16, pp. 1–29, 1999.
- [56] R. P. Abelson. Mathematical models of the distribution of attitudes under controversy. *Contributions to Mathematical Psychology*, N. Fredericksen, H. Gulliksen, Eds., New York, USA: Holt, Rinehart and Winston, pp. 1–160, 1964.
- [57] M. Taylor. Towards a mathematical theory of influence and attitude change. *Human Relations*, vol. 21, no. 2, pp. 121–139, 1968. DOI: 10.1177/001872676802100202.
- [58] Z. H. Yang, Y. Song, M. Zheng, W. Y. Hou. Consensus of multi-agent systems under switching agent dynamics and jumping network topologies. *International Journal of Automation and Computing*, vol. 13, no. 5, pp. 438–446, 2016. DOI: 10.1007/s11633-016-0960-z.
- [59] H. J. Savino, F. O. Souza, L. C. A. Pimenta. Consensus on intervals of communication delay. *International Journal of Automation and Computing*, vol. 15, no. 1, pp. 13–24, 2018. DOI: 10.1007/s11633-017-1095-6.
- [60] Z. Xu, J. Liu, T. Başar. On a modified DeGroot-Friedkin model of opinion dynamics. In *Proceedings of American Control Conference*, IEEE, Chicago, USA, pp. 1047–1052, 2015. DOI: 10.1109/ACC.2015.7170871.
- [61] W. G. Xia, J. Liu, K. H. Johansson, T. Başar. Convergence rate of the modified DeGroot-Friedkin model with doubly stochastic relative interaction matrices. In *Proceedings of American Control Conference*, IEEE, Boston, USA, pp. 1054–1059, 2016. DOI: 10.1109/ACC.2016.7525054.
- [62] X. D. Chen, J. Liu, M. A. Belabbas, Z. Xu, T. Başar. Distributed evaluation and convergence of self-appraisals in social networks. *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 291–304, 2017. DOI: 10.1109/TAC.2016.2554280.
- [63] G. Chen, X. M. Duan, N. E. Friedkin, F. Bullo. Social power dynamics over switching and stochastic influence networks. *IEEE Transactions on Automatic Control*, published online. DOI: 10.1109/TAC.2018.2822182.
- [64] M. Ye, B. D. O. Anderson. Effects of behaviour in self-appraisal dynamics on social networks. In *Proceedings of Effects of behaviour in self-appraisal dynamics on social networks*, European Control Conference, Naples, Italy, submitted for publication 2019.
- [65] N. E. Friedkin. A formal theory of reflected appraisals in the evolution of power. *Administrative Science Quarterly*, vol. 56, no. 4, pp. 501–529, 2011. DOI: 10.1177/0001839212441349.
- [66] A. Mirtabatabaei, P. Jia, N. E. Friedkin, F. Bullo. On the reflected appraisals dynamics of influence networks with stubborn agents. In *Proceedings of American Control Conference*, IEEE, Portland, USA, pp. 3978–3983, 2014. DOI: 10.1109/ACC.2014.6859256.
- [67] L. Coch, J. R. Jr. French. Overcoming resistance to change. *Human Relations*, vol. 1, no. 4, pp. 512–532, 1948. DOI: 10.1177/001872674800100408.
- [68] F. M. Thrasher. *The Gang: A Study of 1, 313 Gangs in Chicago*. Chicago, USA: University of Chicago Press, 1963.
- [69] K. Abbink, L. Gangadharan, T. Handfield, J. Thrasher. Peer punishment promotes enforcement of bad social norms. *Nature Communications*, vol. 8, Article number 609, 2017. DOI: 10.1038/s41467-017-00731-0.
- [70] N. L. Waters, V. P. Hans. A jury of one: Opinion formation, conformity, and dissent on juries. *Journal of Empirical Legal Studies*, vol. 6, no. 3, pp. 513–540, 2009. DOI: 10.1111/j.1740-1461.2009.01152.x.
- [71] R. Willer, K. Kuwabara, M. W. Macy. The false enforcement of unpopular norms. *American Journal of Sociology*, vol. 115, no. 2, pp. 451–490, 2009. DOI: 10.1086/599250.
- [72] H. J. O’Gorman. Pluralistic ignorance and white estimates of white support for racial segregation. *Public Opinion Quarterly*, vol. 39, no. 3, pp. 313–330, 1975. DOI: 10.1086/268231.
- [73] F. H. Allport. *Social Psychology*, Boston, USA: Houghton Mifflin Company, 1924.
- [74] S. Tanford, S. Penrod. Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, vol. 95, no. 2, pp. 189–225, 1984. DOI: 10.1037/0033-2909.95.2.189.
- [75] G. Stasser, J. H. Davis. Group decision making and social influence: A social interaction sequence model. *Psychological Review*, vol. 88, no. 6, pp. 523–551, 1981. DOI: 10.1037/0033-295X.88.6.523.
- [76] B. Mullen. Operationalizing the effect of the group on the individual: A self-attention perspective. *Journal of Experimental Social Psychology*, vol. 19, no. 4, pp. 295–322, 1983. DOI: 10.1016/0022-1031(83)90025-2.
- [77] R. Bond. Group size and conformity. *Group Processes & Intergroup Relations*, vol. 8, no. 4, pp. 331–354, 2005. DOI: 10.1177/1368430205056464.
- [78] D. A. Prentice, D. T. Miller. Pluralistic ignorance and the perpetuation of social norms by unwitting actors. *Advances in Experimental Social Psychology*, vol. 28, pp. 161–209, 1996. DOI: 10.1016/S0065-2601(08)60238-5.
- [79] M. Ye, Y. Z. Qin, A. Govaert, B. D. O. Anderson, M. Cao. *An influence network model to study discrepancies in expressed and private opinions*, [Online], Available: <https://arxiv.org/abs/1806.11236>, 2018.
- [80] E. Noelle-Neumann. *The Spiral of Silence: Public Opinion-Our Social Skin*, Chicago, USA: University of Chicago Press, 1993.
- [81] M. Ye, M. H. Trinh, Y. H. Lim, B. D. O. Anderson, H. S. Ahn. *Continuous-time opinion dynamics on multiple interdependent topics*, [Online], Available: <https://arxiv.org/abs/1805.02836>, 2018.
- [82] T. Kuran. Sparks and prairie fires: A theory of unanticipated political revolution. *Public Choice*, vol. 61, no. 1, pp. 41–74, 1989. DOI: 10.1007/BF00116762.
- [83] J. Goodwin. Why we were surprised (Again) by the Arab spring. *Swiss Political Science Review*, vol. 17, no. 4, pp. 452–456, 2011. DOI: 10.1111/j.1662-6370.2011.02045.x.
- [84] D. G. Taylor. Pluralistic ignorance and the spiral of silence: A formal analysis. *Public Opinion Quarterly*, vol. 46, no. 3, pp. 311–335, 1982. DOI: 10.1086/268729.
- [85] A. N. Awan. Virtual jihadist media: Function, legitimacy and radicalizing efficacy. *European Journal of Cultural Studies*, vol. 10, no. 3, pp. 389–408, 2007. DOI: 10.1177/1367549407079713.



- [86] M. Ye, J. Liu, B. D. O. Anderson. On the effects of heterogeneous logical interdependencies in multi-dimensional opinion dynamics models. In *Proceedings of the 57th IEEE Conference on Decision and Control*, Miami Beach, USA, pp. 4372–4377, 2018. DOI: 10.1109/CDC.2018.8619674.
- [87] M. J. Eppler, J. Mengis. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, vol. 20, no. 5, pp. 325–344, 2004. DOI: 10.1080/01972240490507974.
- [88] J. R. Fox, B. Park, A. Lang. When available resources become negative resources: The effects of cognitive overload on memory sensitivity and criterion bias. *Communication Research*, vol. 34, no. 3, pp. 277–296, 2007. DOI: 10.1177/0093650207300429.



**Brian D. O. Anderson** received the B.Sc. degree in pure mathematics in 1962, and B.Eng. degree in electrical engineering in 1964, from the University of Sydney, Australia, and the Ph.D. degree in electrical engineering from Stanford University, USA in 1966. He is an emeritus professor at the Australian National University (having retired as distinguished professor in 2016), distinguished professor at Hangzhou Dianzi University, and distinguished researcher in Data61-CSIRO, Australia. His awards include the IEEE Control Systems Award of 1997, the 2001 IEEE James H Mulligan, Jr Education Medal,

and the Bode Prize of the IEEE Control System Society in 1992, as well as several IEEE and other best paper prizes. He is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Royal Society, and a foreign member of the US National Academy of Engineering. He holds honorary doctorates from a number of universities, including Université Catholique de Louvain, Belgium, and Eidgenoessische Technische Hochschule (Swiss Federal Institute of Technology), Zurich. He is a past president of the International Federation of Automatic Control and the Australian Academy of Science.

His research interests include distributed control and econometric modelling.

E-mail: Brian.Anderson@anu.edu.au (Corresponding author)  
ORCID iD: 0000-0002-1493-4774



**Mengbin Ye** received the B.Eng. degree (with First Class Honours) in mechanical engineering from University of Auckland, New Zealand in 2013, and the Ph.D. degree in engineering at the Australian National University, Australia in 2018. He is currently a postdoctoral researcher with the Faculty of Science and Engineering, University of Groningen, the Netherlands.

His research interests include opinion dynamics and social networks, consensus and synchronisation of Euler-Lagrange systems, and localisation using bearing measurements.

E-mail: m.ye@rug.nl  
ORCID iD: 0000-0003-1698-0173